Individualized optimal decision making, an artificial intelligence paradigm tailored to an individual's characteristics, has attracted growing attention in many fields. Examples include developing an individualized treatment rule for patients to optimize expected clinical outcomes of interest in medicine, offering customized incentives to increase sales and level of engagement in economics, and designing a personalized advertisement recommendation system to raise the click rates in marketing. Among all individualized decision rules (IDR) that assign treatments to subjects based on their observed covariates, the one that maximizes the expected outcome is referred to as an optimal IDR. Though there is a huge literature on learning the optimal IDR, most of the existing works focus on simple settings, i.e., the discrete treatment domain, a single dataset, and complete data structure, making these methods less practical in reality. On the other hand, prior to adopting any decision rule in practice, it is also crucial to know the impact of implementing such a policy, and thus, evaluating the decision rule offline attracts a lot of attention by using observational data. Despite the popularity of developing methods for policy optimization and evaluation with simple settings such as discrete treatments, less attention has been paid to complex data.

In this thesis, we mainly focus on addressing individualized optimal decision making and policy evaluation with complex data from real-world applications in precision medicine, customized economics, personalized marketing, modern epidemiology, etc. This thesis is structured as follows. In Chapters 2-3, we consider policy evaluation and optimal treatment estimation with continuous treatments. Specifically, in Chapter 2, to handle continuous treatment settings such as personalized dose-finding and dynamic pricing, we propose a jump interval-learning to develop an individualized interval-valued decision rule (I2DR) that maximizes the expected outcome. Unlike IDRs that recommend a single treatment, the proposed I2DR yields an interval of treatment options for each individual, making it more flexible to implement in practice. Statistical properties of the resulting I2DR are established, followed by a procedure to infer its mean outcome. Another goal in precision medicine or customized economics is to evaluate the impact of running a new treatment/pricing strategy based on the historical data that is generated by a different strategy. This enables the clinicians or agents to evaluate the performance of different IDRs that are not employed
but of great interest. To handle continuous treatments, in Chapter 3, we develop a novel estimation method for off-policy evaluation using deep jump learning. The key ingredient of our method lies in adaptively discretizing the treatment space using deep discretization, by leveraging deep learning and multi-scale change point detection. This allows us to apply existing off-policy evaluation methods in discrete treatments to handle continuous treatments, and overcome the limitations of kernel-based methods.

In Chapters 4-5, we propose methods to estimate IDR for other types of complex data, including multiple datasets from heterogeneous studies and incomplete data structures. In Chapter 4, we consider estimating the optimal IDR in a primary sample of interest with multiple auxiliary sources available, as widely observed in the electronic health/medical records. The outcome of interest is limited in the sense that it is only observed in the primary sample. In reality, these multiple data sources may belong to heterogeneous studies and thus cannot be combined directly. We propose a new framework to handle heterogeneous studies and address the limited outcome simultaneously through a novel calibrated optimal decision-making method, by leveraging the common intermediate outcomes in multiple data sources. Our method allows the baseline covariates across different samples to have either homogeneous or heterogeneous distributions. The proposed estimator is shown to be asymptotically normal and more efficient than using the primary sample solely. A more challenging case occurs when treatments have a long-term effect, and as such, the outcome of interest cannot be observed as well in the primary sample due to the limited duration of experiments, which makes the estimation of optimal IDR impossible. In Chapter 5, we address this challenge by making use of the auxiliary sample to facilitate the estimation of optimal IDR in the primary sample via specifying the missing mechanism.
On Optimal Decision Making and Policy Evaluation with Complex Data

by
Hengrui Cai

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina
2022

APPROVED BY:

________________________     ________________________
Marie Davidian                Shu Yang

________________________     ________________________
Wenbin Lu                    Rui Song
Co-chair of Advisory Committee Co-chair of Advisory Committee
DEDICATION

To my mothers and grandmother.
BIOGRAPHY

The author was born in Ruian, China two days before Moon Festival in 1994. She graduated from Zhejiang Ruian High School in 2013 and attended Zhejiang University afterward. After receiving her bachelor’s degree in Statistics and Advanced Honor Class of Engineering Education in Chu Kochen Honors College in 2017, she continued her study in the Department of Statistics at North Carolina State University. Co-advised by Dr. Wenbin Lu and Dr. Rui Song, she will complete her Ph.D. in Statistics in 2022 May.
ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my co-advisors, Dr. Wenbin Lu and Dr. Rui Song, for their excellent guidance and generous support. During the past five years, I have received great help from them to lead me to front-end statistical research and a cheerful Ph.D. career. Starting in my first year, they gave me numerous suggestions on selecting courses as well as attending insightful seminars every semester, which vastly enrich my knowledge and broaden my horizons across the statistical world. The thing I most look forward to is the weekly meeting with Dr. Lu and Dr. Song, where we brainstorm ideas for stubborn problems, examine methods for perfect results, and refresh our eyes for future works. I learned a lot from them to always keep a creative eye and critical thinking towards the world. Their enthusiasm towards science and excellent capability as the domain experts immensely support my research works. Also, I appreciate their time and own experience they shared with me, which is very valuable to me in deciding my career path and strengthens my faith in academia. Under the supervision of Dr. Song and Dr. Lu, I have completed over 10 projects with 7 published at journals and conferences. As of Fall 2022, I will be an Assistant Professor in the Department of Statistics at the University of California Irvine. I was very fortunate to work with them during my Ph.D. Without their selfless support, I cannot achieve what I have done so far.

I would also like to sincerely thank my committee members, Dr. Marie Davidian, Dr. Yang Shu, and Dr. Sanmugavadivel Ranjithan, for their valuable advice and great help over the years. Dr. Marie Davidian taught one of my favorite classes 'Dynamic Treatment Regimes'. This course completely changed my thoughts on investigating causal relationships and stimulated my interest in estimation and inference of optimal treatment regimes, which is the focus of this thesis. I am also grateful for the tremendous and timely support and valuable suggestions I received from Dr. Davidian during my application season for faculty jobs. I am also very fortunate to have Dr. Yang Shu as my committee member, who has provided so much constructive feedback on my research work and gave me a very good example as a young outstanding researcher. Finally, I would like to sincerely thank Dr. Sanmugavadivel Ranjithan for offering to be my graduate school representative and for his valuable comments on my research works.

It was my great pleasure and honor to collaborate with many talented researchers in these five years, including Drs. Chengchun Shi, Ye Shen, Runzhe Wan, Zhihao Cen, Richard Watson, Xinming An, Haoyu Wei, Dale Gao, Sheng Zhang, Jianian Wang, Ling Leng, Roman
Levkin, Chirag Mandaviya, Rachel Marceau West, Devan V. Mehrotra, Lingkang Huang, etc. Their insightful comments and suggestions enormously helped me in my research. I also would like to express my thanks to Dr. Dennis Boos, Dr. Rui Song, Dr. Jacqueline M. Hughes-Oliver, Dr. Daowen Zhang, Dr. Tom Reiland, Dr. Emily Griffith, Dr. Emily Hector, Dr. Nathan Corder, and Dr. Tyler Johnson. I worked with them as a teaching assistant and learned a lot from them. I would like to thank Dr. Dennis Boos and Dr. Emily Hector for their great support during my application season. I am also grateful to all the professors in the Department of Statistics at North Carolina State University for offering excellent courses on various topics that lay the foundation and broaden my knowledge of statistics, including Drs. Butch Tsiatis, Marie Davidian, Wenbin Lu, Subhashis Ghoshal, Sujit Ghosh, Jacqueline M. Hughes-Oliver, Ana-Maria Staicu, Ryan Martin, Arnab Maity, Eric Chi, Emily Griffith, and many others. I also appreciate the department staff, Lanakila, Terry, Alison, Kathleen, Sasha, and Melissa, for their generous help and warm reminders during my Ph.D. life.

I felt so lucky to have all the talented and lovely friends, classmates, office mates, and roommates in the past five years in Raleigh. I would like to thank my friends, Yifei Hu, Yiming Wang, Runzhe Wan, Lili Wu, Rui Zhu, Lu Lu, Yuan Tian, Chengchun Shi, Yuan Feng, Qi Ma, Xuan Liu, Wei Zhao, Alex Chen, and many others. They shared laughs and tears with me over the years and made my graduate school much more colorful. Particular thanks go to my roommate, Ruonan Li, for her optimistic attitude towards life and super warm company during the pandemic. Lastly, I would also thank my best friend Jiayi Wang, who always comforted and encouraged me whenever I felt frustrated. I owe too many people and it is impossible to express my all thanks on this short page.

Most importantly, I would thank my family. My mothers, Aixia Wu and Lvmei Lin, have always been there to support and love me unconditionally. I am also grateful to my nana, Ajuan Ruan, for raising me up and leading me to be an independent woman. They always encourage me to chase to do whatever I want. My gratitude for what they have done goes far beyond the words. I have nothing but the best wishes for them to be healthy and happy.
# TABLE OF CONTENTS

List of Tables ........................................................................................................... x

List of Figures ............................................................................................................ xii

Chapter 1 INTRODUCTION .................................................................................. 1

1.1 Individualized Decision Making and Policy Evaluation ................................ 1

1.2 Challenges in Policy Optimization and Evaluation with Complex Data ........ 2

1.2.1 Continuous Treatment Domains .............................................................. 2

1.2.2 Multiple Data Sources and Incomplete Data Structure .......................... 4

1.3 Outline .............................................................................................................. 5

Chapter 2 Jump Interval-Learning for Individualized Decision Making ............... 7

2.1 Introduction ...................................................................................................... 7

2.2 Statistical Framework .................................................................................... 10

2.2.1 Model Setup ............................................................................................ 10

2.2.2 I2DR ......................................................................................................... 11

2.2.3 Working Model Assumptions .................................................................. 12

2.3 Methods ......................................................................................................... 12

2.3.1 Jump Interval-learning ............................................................................ 12

2.3.2 Linear- and Deep-JIL ............................................................................. 15

2.3.3 Implementation ....................................................................................... 16

2.3.4 Tuning Parameters .................................................................................. 19

2.4 Theory ............................................................................................................ 20

2.4.1 Properties under Model I ....................................................................... 21

2.4.2 Properties under Model II ...................................................................... 26

2.5 Simulations ..................................................................................................... 28

2.5.1 Confidence Interval for the Value ......................................................... 28

2.5.2 Value Function under the Proposed I2DR ........................................... 29

2.5.3 Choice of \( m \) ......................................................................................... 32

2.6 Real Data Analysis .......................................................................................... 33

2.7 Discussions ..................................................................................................... 35

2.7.1 Diverging Number of Change Points .................................................... 35

2.7.2 Potential Alternative Approaches .......................................................... 36

2.7.3 Other Penalty Functions in L-JIL .......................................................... 38

Chapter 3 Deep Jump Learning for Off-Policy Evaluation in Continuous Treatment Settings .................................................................................................................. 39

3.1 Introduction ...................................................................................................... 39

3.2 Preliminaries ................................................................................................... 40

3.2.1 Problem Formulation .............................................................................. 40

3.2.2 Doubly Robust Estimator and Kernel-Based Evaluation ....................... 41
Chapter 4  Calibrated Optimal Decision Making with Multiple Data Sources and Limited Outcome 58
4.1 Introduction 58
4.2 Statistical Framework 59
4.2.1 Notation and Formulation 59
4.2.2 Assumptions 61
4.2.3 Related Works 62
4.3 Method 63
4.3.1 Calibrated Optimal Decision Making for Homogeneous Baseline Covariates 63
4.3.2 Calibrated Optimal Decision Making for Heterogeneous Baseline Covariates 66
4.3.3 Iterative Policy Tree Search Algorithm 69
4.4 Theoretical Properties 70
4.5 Simulation Studies 73
4.5.1 Evaluation on Calibrated Value Estimator for Homogeneous Baseline Covariates 73
4.5.2 Evaluation on Calibrated Value Estimator for Heterogeneous Baseline Covariates 76
4.6 Real Data Analysis 78
4.7 Discussions 80

Chapter 5  On Optimal Decision Making with Auxiliary Data for Long-Term Effect 82
5.1 Introduction 82
5.2 Statistical Framework 84
5.2.1 Experimental Sample and Auxiliary Sample 84
5.2.2 Assumptions 85
5.2.3 Value Function and Optimal Decision Rule 87
5.3 Proposed Method 87


<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.1 Proposed Estimator for Long-Term Outcome</td>
<td>87</td>
</tr>
<tr>
<td>5.3.2 Implementation Details</td>
<td>89</td>
</tr>
<tr>
<td>5.3.3 Theoretical Properties</td>
<td>89</td>
</tr>
<tr>
<td>5.4 Simulation Studies</td>
<td>93</td>
</tr>
<tr>
<td>5.4.1 Evaluation under Correctly Specified Model</td>
<td>93</td>
</tr>
<tr>
<td>5.4.2 Evaluation under Model Misspecification</td>
<td>95</td>
</tr>
<tr>
<td>5.5 Real Data Analysis</td>
<td>98</td>
</tr>
<tr>
<td>5.6 Discussions</td>
<td>99</td>
</tr>
<tr>
<td>References</td>
<td>100</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>110</td>
</tr>
<tr>
<td>Appendix A SUPPLEMENT TO CHAPTER 2</td>
<td>111</td>
</tr>
<tr>
<td>A.1 More on Tuning in L-JIL</td>
<td>111</td>
</tr>
<tr>
<td>A.2 Technical Proofs</td>
<td>112</td>
</tr>
<tr>
<td>A.2.1 Proof of Theorem 1</td>
<td>113</td>
</tr>
<tr>
<td>A.2.2 Proof of Lemma 1</td>
<td>125</td>
</tr>
<tr>
<td>A.2.3 Proof of Lemma 2</td>
<td>132</td>
</tr>
<tr>
<td>A.2.4 Proof of Lemma 3</td>
<td>134</td>
</tr>
<tr>
<td>A.2.5 Proof of Lemma 4</td>
<td>137</td>
</tr>
<tr>
<td>A.2.6 Proof of Theorem 2</td>
<td>140</td>
</tr>
<tr>
<td>A.2.7 Proof of Theorem 3</td>
<td>148</td>
</tr>
<tr>
<td>A.2.8 Proof of Lemma 5</td>
<td>152</td>
</tr>
<tr>
<td>A.2.9 Proof of Lemma 6</td>
<td>157</td>
</tr>
<tr>
<td>A.2.10 Proof of Lemma 7</td>
<td>158</td>
</tr>
<tr>
<td>A.2.11 Proof of Theorem 4</td>
<td>158</td>
</tr>
<tr>
<td>A.2.12 Proof of Theorem 5</td>
<td>162</td>
</tr>
<tr>
<td>A.2.13 Proof of Theorem 6</td>
<td>166</td>
</tr>
<tr>
<td>A.2.14 Proof of Lemma 8</td>
<td>177</td>
</tr>
<tr>
<td>A.2.15 Proof of Theorem 7</td>
<td>180</td>
</tr>
<tr>
<td>Appendix B SUPPLEMENT TO CHAPTER 3</td>
<td>186</td>
</tr>
<tr>
<td>B.1 Analysis of Computational Complexity of DJL</td>
<td>186</td>
</tr>
<tr>
<td>B.2 Additional Experimental Results</td>
<td>188</td>
</tr>
<tr>
<td>B.3 Rate of Convergence of Kernel-Based Estimators</td>
<td>191</td>
</tr>
<tr>
<td>B.3.1 Convergence Rate under Model 1</td>
<td>191</td>
</tr>
<tr>
<td>B.3.2 Convergence Rate under Model 2</td>
<td>192</td>
</tr>
<tr>
<td>B.4 Technical Proofs</td>
<td>193</td>
</tr>
<tr>
<td>B.4.1 Proof of Lemma B.4.1</td>
<td>194</td>
</tr>
<tr>
<td>B.4.2 Proof of Lemma B.4.2</td>
<td>200</td>
</tr>
<tr>
<td>B.4.3 Proof of Lemma B.4.3</td>
<td>200</td>
</tr>
<tr>
<td>B.4.4 Proof of Theorem 3.4.1</td>
<td>201</td>
</tr>
<tr>
<td>B.4.5 Proof of Theorem 3.4.2</td>
<td>209</td>
</tr>
<tr>
<td>Appendix C SUPPLEMENT TO CHAPTER 4</td>
<td>214</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>C.1 Estimation on Variances</td>
<td>214</td>
</tr>
<tr>
<td>C.2 Extension of Iterative Policy Tree Search Algorithm</td>
<td>216</td>
</tr>
<tr>
<td>C.3 Investigation with Multiple Intermediate Outcomes</td>
<td>217</td>
</tr>
<tr>
<td>C.4 Technical Proofs</td>
<td>219</td>
</tr>
<tr>
<td>C.4.1 Proof of Lemma 4.3.1</td>
<td>219</td>
</tr>
<tr>
<td>C.4.2 Proof of Lemma 4.3.2</td>
<td>220</td>
</tr>
<tr>
<td>C.4.3 Proof of Lemma 4.3.3</td>
<td>222</td>
</tr>
<tr>
<td>C.4.4 Proof of Lemma 4.3.4</td>
<td>225</td>
</tr>
<tr>
<td>C.4.5 Proof of Theorem 4.4.1</td>
<td>226</td>
</tr>
<tr>
<td>C.4.6 Proof of Theorem 4.4.2</td>
<td>232</td>
</tr>
<tr>
<td>C.4.7 Proof of Theorem 4.4.3</td>
<td>237</td>
</tr>
<tr>
<td>Appendix D SUPPLEMENT TO CHAPTER 5</td>
<td>238</td>
</tr>
<tr>
<td>D.1 Inverse Propensity-Score Weighted Estimator</td>
<td>238</td>
</tr>
<tr>
<td>D.1.1 IPW Estimator for the Long-term Outcome</td>
<td>238</td>
</tr>
<tr>
<td>D.1.2 Theoretical Results of the IPW Estimator</td>
<td>239</td>
</tr>
<tr>
<td>D.2 Technical Proofs</td>
<td>240</td>
</tr>
<tr>
<td>D.2.1 Proof of Lemma 5.3.1</td>
<td>240</td>
</tr>
<tr>
<td>D.2.2 Proof of Lemma 5.3.2</td>
<td>242</td>
</tr>
<tr>
<td>D.2.3 Proof of Theorem D.1.1</td>
<td>242</td>
</tr>
<tr>
<td>D.2.4 Proof of Lemma D.1.1 and Lemma 5.3.3</td>
<td>245</td>
</tr>
<tr>
<td>D.2.5 Proof of Theorem D.1.2</td>
<td>250</td>
</tr>
<tr>
<td>D.2.6 Proof of Theorem 5.3.2</td>
<td>252</td>
</tr>
<tr>
<td>D.3 Sensitivity Studies</td>
<td>254</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1 The estimated optimal value \( \hat{V} \) with its standard error, the empirical coverage probability of its associated confidence interval, and the averaged number of estimated partitions computed by the proposed L-JIL and D-JIL. ................................................................. 29
Table 2.2 The value function under the proposed I2DR and DRs estimated based on outcome weighted learning for Scenario 1-5. ................. 31
Table 2.3 The value function of the proposed I2DR under L-JIL for Scenario 1-5 with different choice of \( m = n/c \). ......................................................... 32
Table 2.4 The value function of the proposed I2DR under D-JIL for Scenario 1-5 with different choice of \( m = n/c \). ......................................................... 33
Table 2.5 The regression coefficients associated with three subintervals for the proposed I2DR under L-JIL. ................................................................. 34
Table 3.1 The absolute error and the standard deviation (in parentheses) of the estimated values for \( V^{(1)} \) and \( V^{(2)} \), using DJL and kernel-based methods, for target policy \( d(x) = x \) in the toy example. ................................. 44
Table 3.2 The absolute error, the standard deviation, and the mean squared error of the estimated values under the optimal policy via different methods for the Warfarin data. The target value is given by \(-0.278\). . 56
Table 4.1 Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenario 1... 75
Table 4.2 Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenario 2, where \( \max_{\Pi_1^{(d)}} V(d) = 1.251 \). ................................................................. 75
Table 4.3 Empirical results of the proposed CODA-HE method in comparison to the ODR based on the primary sample solely under Scenario 1 with heterogeneous baseline covariates. ............................. 77
Table 4.4 Empirical results of the proposed CODA-HE method in comparison to the ODR based on the primary sample solely under Scenario 2 with heterogeneous baseline covariates, where \( \max_{\Pi_1^{(d)}} V(d) = 1.251 \). . 77
Table 4.5 The real data analysis under the proposed CODA method and the ODR method based on the primary sample solely. ............................... 80
Table 5.1 Empirical results under the GEAR for Scenario 1 and 2. ............. 95
Table 5.2 Empirical results under the GEAR for Scenario 3-5. .................. 97
Table 5.3 Comparison results for ACTG 175 data. ................................. 99
Table B.1 The averaged computational cost (in minutes) under the proposed deep jump learning and three kernel-based methods for Scenario 1. . 188
Table B.2 The averaged computational cost under the proposed deep jump learning for Scenario 1 with large sample settings. .................................. 189

Table B.3 The absolute error and the standard deviation (in parentheses) of the estimated values under the optimal policy via the proposed deep jump learning and three kernel-based methods for Scenario 1 to 4. .............. 189

Table B.4 The averaged size of the final estimated partition (|Ê|) in comparison to the initial number of intervals (m) under the proposed DJL for Scenario 1 to 4. ................................................................. 190

Table B.5 The mean squared error (MSE), the normalized root-mean-square-deviation (NRMSD), the mean absolute error (MAE), and the normalized MAE (NMAE) of the fitted model under the multilayer perceptrons regressor, linear regression, and the random forest algorithm, via ten-fold cross-validation. ........................................ 191

Table C.1 Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenarios 3 to 5.218

Table D.1 Empirical results of ˆβG and its estimated value ˆVAP(ˆβG) under the GEAR for Scenario 6 when l = {0, 0.4, 0.8}. Note the true parameter of the ODR is β0 = [0, 1/√2, −1/√2]T with the true value 0.333. .......... 255
LIST OF FIGURES

Figure 2.1 Illustration of DNN with $L = 2$ and $W = 25$; here $\mu \in \mathbb{R}^p$ is the input, the output is given by $A^{(3)}\sigma(A^{(2)}\sigma(A^{(1)}\mu + b^{(1)}) + b^{(2)}) + b^{(3)}$ where $A^{(l)}$, $b^{(l)}$ denote the corresponding parameters to produce the linear transformation for the $(l-1)$th layer and that $\sigma$ denotes the component-wise ReLU function. In this example, $W = \sum_{j=1}^{3}(\|A^{(3)}\|_0 + \|b^{(3)}\|_0) = 25$ where $\|\cdot\|_0$ denotes the number of nonzero elements in the vector or matrix. ......................................................... 17

Figure 3.1 Left panel: example of piece-wise constant function. Middle panel: the oracle conditional mean function $Q$ on the feature-treatment space for the toy example. Right panel: the green curve presents the oracle $Q\{x, d(x)\}$ under target policy $d(x) = x$ in the toy example; and the red curve is the fitted mean value by DJL and the pink dash line corresponds to the 95% confidence bound. ................................. 44

Figure 3.2 The box plot of the estimated values under the optimal policy via the proposed DJL method and three kernel-based methods for Scenario 1-4. The target values are 1.33, 1, 4.86 and 1.6, respectively. ............... 54

Figure 4.1 The box-plots for the shared baseline variables in the MIMIC-III data and the eICU data. ................................................................. 79

Figure 5.1 A direct acyclic graph illustrating assumptions (A2), (A4), and (A5) in the joint sample. White nodes represent observed variables, and grey nodes are variables with missing values. ........................................ 86

Figure B.1 The absolute error of the estimated value and the computational cost (in minutes) under the DJL with different initial number of intervals ($m$) when $n = 100$ in Scenario 1. ................................. 190

Figure D.1 The trend of the bias of $V(\hat{\beta}^G)$ under the GEAR over the parameter $l$. 256

Figure D.2 The trend of the average rate of the correct decision made by the GEAR over the parameter $l$. ............................................. 256
1.1 Individualized Decision Making and Policy Evaluation

Individualized decision making is an increasingly attractive artificial intelligence paradigm that proposes to assign each individual a given treatment based on their observed characteristics, with a wide variety of applications. Examples include developing an individualized treatment rule for patients to optimize expected clinical outcomes of interest in precision medicine (Chakraborty and Moodie 2013; Collins and Varmus 2015), offering customized incentives to increase sales and level of engagement in economics (Qiang and Bayati 2016; Turvey 2017), and designing a personalized advertisement recommendation system to raise the click rates in marketing (McInerney et al. 2018; Fong et al. 2018), etc. The resulting decision strategy is referred to as an individualized decision rule (IDR) that assign treatments to subjects based on their observed covariates. Among all individualized decision rules (IDR), the one that maximizes the expected outcome is referred to as an optimal IDR. There is a huge literature on learning the optimal IDR from the observed dataset. Some popular methods include Q-learning (Watkins and Dayan 1992; Chakraborty et al. 2010; Qian and Murphy 2011; Song et al. 2015), A-learning (Robins 2004; Murphy 2003a; Shi et al.
2018), policy search methods (Zhang et al. 2012a, 2013a; Wang et al. 2018; Nie et al. 2020), outcome weighted learning (Zhao et al. 2012a, 2015; Zhu et al. 2017; Meng et al. 2020), concordance-assisted learning Fan et al. (2017); Liang et al. (2017), decision list-based methods (Zhang et al. 2015, 2018), and direct learning (Qi et al. 2020).

On the other hand, prior to adopting any decision rule in practice, it is crucial to know the impact of implementing such a rule. In medical and public-policy domains, it is risky to apply a decision rule or policy online to estimate its mean outcome (see, e.g., Murphy et al. 2001; Hirano et al. 2003). Policy evaluation thus attracts a lot of attention, by making use of the offline data that was generated by a different historical decision rule. Another goal in precision medicine or customized economics is to unbiasedly estimate the mean outcome under a new decision rule based on such a pre-collected dataset. The conditional mean of the outcome under a decision rule is also known as the value function in the literature of individualized treatment regime. Despite the popularity of evaluating decision rules offline with a finite set of treatment options, (see e.g., Wang et al. 2012; Zhang et al. 2012b, 2013a; Dudík et al. 2014; Chakraborty et al. 2014; Matsouaka et al. 2014; Luedtke and Van Der Laan 2016; Shi et al. 2020), less attention has been paid to complex data.

1.2 Challenges in Policy Optimization and Evaluation with Complex Data

Though there is a huge literature on learning the optimal IDR, We note however, most of the existing works focus on simple settings, i.e., the discrete treatment domain where the number of available treatment options is finite, a single dataset, and complete data structure, making these methods less practical in reality. In this section, we introduce several challenges in policy Optimization and evaluation with complex data, including continuous treatment domains, multiple datasets from heterogeneous studies, and incomplete data structures.

1.2.1 Continuous Treatment Domains

Recently, more attention has been paid on individualized decision making with continuous treatment domains. These studies occur in a number of real applications, including personalized dose finding (Chen et al. 2016) and dynamic pricing (den Boer and Keskin 2020). For instance, in personalized dose finding, one wishes to derive a dose level or dose
range for each patient. Due to patients' heterogeneity in response to doses, it is commonly assumed that there may not exist a unified best dose for all patients. Thus, one major interest in precision medicine is to develop an IDR that assigns each individual patient with a certain dose level or a specified range of doses from a continuous domain based on their individual personal information, to optimize their health status. Similarly, in dynamic pricing, we aim to identify an IDR that assigns each product an optimal price according to their characteristics to maximize the overall profit.

In contrast to developing the optimal IDR under discrete treatment settings, individualized decision making with a continuous treatment domain has been less studied. Among those available, Rich et al. (2014) modeled the interactions between the dose level and covariates to recommend personalized dosing strategies. Laber and Zhao (2015) developed a tree-based method to derive the IDR by dividing patients into subgroups and assigning each subgroup the same dose level. Chen et al. (2016) proposed an outcome weighted learning method to directly search the optimal IDR among a restricted class of IDRs. Zhou et al. (2018) proposed a dimension reduction framework to personalized dose finding that effectively reduces the dimensionality of baseline characteristics from high to a moderate scale. Kallus and Zhou (2018) and Chernozhukov et al. (2019) evaluated and optimized IDRs for continuous treatments by replacing the indicator function in the doubly-robust approach with the kernel function, and by modeling the conditional mean outcome function (i.e., the value) through a semi-parametric form, respectively. Zhu et al. (2020a) focused on the class of linear IDRs and proposed to compute an optimal linear IDR by maximizing a kernel-based value estimate. Schulz and Moodie (2020) proposed a doubly robust estimation method for personalized dose finding. The estimated optimal IDRs computed by these methods typically recommend one single treatment level for each individual, making it hard to implement in practice.

Similarly, though there is a the popularity of developing OPE methods with finitely many treatment (or action) options (see e.g., Dudík et al. 2011, 2014; Wang et al. 2012; Zhang et al. 2012b, 2013b; Luedtke and Van Der Laan 2016; Jiang and Li 2016; Swaminathan et al. 2017; Wang et al. 2017; Farajtabar et al. 2018; Cai et al. 2020; Wu and Wang 2020; Su et al. 2020a; Kallus and Uehara 2020a; Shi et al. 2020, 2021), less attention has been paid to the continuous treatment setting, such as personalized dose finding (Chen et al. 2016; Zhou et al. 2018; Zhu et al. 2020a,b), dynamic pricing (den Boer and Keskin 2020), and contextual bandits (Chernozhukov et al. 2019). Recently, a few OPE methods have been proposed to handle continuous treatments (Kallus and Zhou 2018; Krishnamurthy et al. 2019; Sondhi et al. 2020; Colangelo and Lee 2020; Singh et al. 2020; Su et al. 2020b;
Kallus and Uehara 2020b). All these methods rely on the use of a kernel function to extend the inverse probability weighting (IPW) or doubly robust (DR) approaches developed in discrete treatment domains (see e.g., Dudík et al. 2011). They suffer from two limitations. First, the validity of these methods requires the mean outcome to be a smooth function over the treatment space. This assumption could be violated in applications such as dynamic pricing, where the expected demand for a product has jump discontinuities as a function of the charged price (den Boer and Keskin 2020). Specifically, a product could attract a new segment of customers if the seller lowers the price below a certain threshold. Thus, there will be a sudden increase in demand by a small price reduction, yielding a discontinuous demand function. Second, these kernel-based methods typically use a single bandwidth parameter. This is sub-optimal in cases where the second-order derivative of the conditional mean function has an abrupt change in the treatment space; see Section 3.3.1 for details. Addressing these limitations requires the development of new policy evaluation tools and theory.

### 1.2.2 Multiple Data Sources and Incomplete Data Structure

Most existing IDR methods are developed based on data from a single source where the primary outcome of interest can be observed for all subjects, making these works less practical in more complicated situations. There are many applications involving multiple datasets from different sources, where the primary outcome of interest is limited in the sense that it is only observed in some data sources. Take the treatment of sepsis as an instance. In the Medical Information Mart for Intensive Care (MIMIC-III) clinical database (Goldberger et al. 2000; Johnson et al. 2016; Bisera et al. 2020), thousands of patients in intensive care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 were treated with different medical supervisions such as the vasopressor and followed up for the mortality due to sepsis as the primary outcome of interest. In addition, we can observe other post-treatment intermediate outcomes (also known as surrogacies or proximal outcomes), such as the total urine output and the cumulated net of metabolism. These intermediate outcomes, as well as baseline variables and the treatment information collected in the MIMIC-III data, were also recorded in the electronic Intensive Care Units (eICU) collaborative research database (Goldberger et al. 2000; Pollard et al. 2018) that contains over 200,000 admissions to intensive care units across the United States between 2014 and 2015. Yet, the outcome of interest was not reported in eICU. Hence, we view the MIMIC-III data as the primary sample and the eICU data as the auxiliary sample without
the outcome of interest, leading to one desideratum in precision medicine on finding the optimal IDR to optimize the mortality rate of sepsis based on these datasets. However, integrating multiple data sources from heterogeneous studies for deriving the optimal IDR can be particularly challenging. Note that the MIMIC-III and eICU data were collected from different locations during different periods. These two samples show certain heterogeneity (see details in Section 4.6) such as diverse probability distributions in baseline covariates, the treatment, and intermediate outcomes, and thus cannot be combined directly. As such, a new optimal IDR estimation approach combining multiple data sources from heterogeneous studies with the limited outcome is desired.

In addition, current optimal IDR methods cannot be applied to cases where treatments have a long-term effect and the primary outcome of interest cannot be observed in the experimental sample. Take the AIDS Clinical Trials Group Protocol 175 (ACTG 175) data (Hammer et al. 1996) as an example. The experiment randomly assigned HIV-infected patients to competitive antiretroviral regimens, and recorded their CD4 count (cells/mm³) and CD8 count over time. A higher CD4 count usually indicates a stronger immune system. However, due to the limitation of the follow-up, the clinical meaningful long-term outcome of interest for the AIDS recovery may be missing for a proportion of patients. Similar problems are also considered in the evaluation of education programs, such as the Student/Teacher Achievement Ratio (STAR) project (Word et al. 1990; Chetty et al. 2011) that studied long-term impacts of early childhood education on the future income. Due to the heterogeneity in individual characteristics, one cannot find a unified best treatment for all subjects. However, the effects of treatment on the long-term outcome of interest cannot be evaluated using the experimental data solely. Hence, deriving an optimal IDR to maximize the expected long-term outcome based on baseline covariates obtained at an early stage is also challenging.

1.3 Outline

In this thesis, we mainly focus on addressing individualized optimal decision making and policy evaluation with complex data from real-world applications in precision medicine, customized economics, personalized marketing, modern epidemiology, etc. This thesis is structured as follows. In Chapters 2-3, we consider policy evaluation and optimal treatment estimation with continuous treatments. Specifically, in Chapter 2, to handle continuous treatment settings such as personalized dose-finding and dynamic pricing, we propose a
jump interval-learning to develop an individualized interval-valued decision rule (I2DR) that maximizes the expected outcome. Unlike IDR{s} that recommend a single treatment, the proposed I2DR yields an interval of treatment options for each individual, making it more flexible to implement in practice. Statistical properties of the resulting I2DR are established, followed by a procedure to infer its mean outcome. Another goal in precision medicine or customized economics is to evaluate the impact of running a new treatment/pricing strategy based on the historical data that is generated by a different strategy. This enables the clinicians or agents to evaluate the performance of different IDR{s} that are not employed but of great interest. To handle continuous treatments, in Chapter 3, we develop a novel estimation method for off-policy evaluation using deep jump learning. The key ingredient of our method lies in adaptively discretizing the treatment space using deep discretization, by leveraging deep learning and multi-scale change point detection. This allows us to apply existing off-policy evaluation methods in discrete treatments to handle continuous treatments, and overcome the limitations of kernel-based methods.

In Chapters 4-5, we propose methods to estimate IDR for other types of complex data, including multiple datasets from heterogeneous studies and incomplete data structures. In Chapter 4, we consider estimating the optimal IDR in a primary sample of interest with multiple auxiliary sources available, as widely observed in the electronic health/medical records. The outcome of interest is limited in the sense that it is only observed in the primary sample. In reality, these multiple data sources may belong to heterogeneous studies and thus cannot be combined directly. We propose a new framework to handle heterogeneous studies and address the limited outcome simultaneously through a novel calibrated optimal decision-making method, by leveraging the common intermediate outcomes in multiple data sources. Our method allows the baseline covariates across different samples to have either homogeneous or heterogeneous distributions. The proposed estimator is shown to be asymptotically normal and more efficient than using the primary sample solely. A more challenging case occurs when treatments have a long-term effect, and as such, the outcome of interest cannot be observed as well in the primary sample due to the limited duration of experiments, which makes the estimation of optimal IDR impossible. In Chapter 5, we address this challenge by making use of the auxiliary sample to facilitate the estimation of optimal IDR in the primary sample via specifying the missing mechanism.
2.1 Introduction

The focus of this chapter is to develop an individualized interval-valued decision rule (I2DR) that returns a range of treatment levels based on individuals’ baseline information. Compared to the IDRs recommended by the existing works, the proposed I2DR gives more options and is thus more flexible to implement in practice. Take personalized dose finding as an illustration. First, interval-valued dose levels may be applied to patients of the same characteristics, when arbitrary dose within the given dose interval could achieve the same efficacy. Studies of the pharmacokinetics of vancomycin conducted by Rotschafer et al. (1982) suggested that adults with normal renal function should receive an initial dosage of 6.5 to 8 milligram of vancomycin per kilogram intravenously over 1 hour every 6 to 12 hours. In the review of warfarin dosing reported by Kuruvilla and Gurk-Turner (2001), when the international normalized ratio (INR) approaches the target range or omit dose, they suggested to give 1-2.5 milligram vitamin K1 if a patient has a risk factor for bleeding, oth-
erwise provide Vitamin K1 2-4 milligram orally. Second, a range of doses gives instructions for designing the medicine specification and helps to save cost on manufacturing dosage. As such, an I2DR is preferred in these applications due to its necessity and flexibility.

Our contributions are summarized as follows. Scientifically, individualized decision making in a continuous treatment domain is a vital problem in many applications such as precision medicine and dynamic pricing. To the best of our knowledge, this is the first work on developing individualized interval-valued decision rules. Our proposal thus fills a crucial gap, extends the scope of existing methods that focus on recommending IDR{s}, and offers a useful tool for individualized decision making in a number of applications.

Methodologically, we propose a novel jump interval-learning (JIL) by integrating personalized decision making with multi-scale change point detection (see Niu et al. 2016, for a selective overview). Our proposal makes useful contributions to the two aforementioned areas simultaneously.

First, to implement personalized decision making, we propose a data-driven I2DR in a continuous treatment domain. Our proposal is motivated by the empirical finding that the expected outcome can be a piecewise function on the treatment domain in various applications. Specifically, in dynamic pricing (Qiang and Bayati 2016; den Boer and Keskin 2020), the expected demand for a product has jump discontinuities as a function of the charged price. This motivates us to impose a piecewise-function model for the expected outcome. We then leverage ideas from the change point detection literature and propose a jump-penalized regression to estimate the conditional mean of the expected outcome as a function of the treatment level and the baseline characteristics (outcome regression function). This partitions the entire treatment space into several subintervals. The proposed I2DR is a set of decision rules that assign each subject to one of these subintervals. In addition, we further develop a procedure to construct a confidence interval (CI) for the expected outcome under the proposed I2DR and the optimal IDR.

Second, we note that most works in the multi-scale change point detection literature either focused on models without covariates, or required the underlying truth to be piecewise constant (see e.g., Boysen et al. 2009; Frick et al. 2014; Fryzlewicz 2014, and the references therein). Our work goes beyond those cited above in that we consider a more complicated (nonparametric) model with covariates, and allow the underlying outcome regression function to be either a piecewise or continuous function over the treatment space. To approximate the expected outcome as a function of baseline covariates, we propose a linear function model and a deep neural networks model. We refer to the two procedures as L-JIL and D-JIL, respectively. Here, the proposed L-JIL yields a set of linear decision rules that
is easy to interpret. See the real data analysis in Section 2.6 for details. On the contrary, the proposed D-JIL employs deep learning (LeCun et al. 2015) to model the complicated outcome-covariates relationships that often occur in high-dimensional settings. We remark that both procedures are developed by imposing a piecewise-function model to approximate the outcome-treatment relationship. Yet, they are valid when the expected outcome is a continuous function of the treatment level as well.

Theoretically, we systematically study the statistical properties of the jump-penalized estimators with linear regression or deep neural networks. Our theoretical approaches can be applied to the analysis of general covariate-based change point models. The model could be either parametric or nonparametric. Specifically, we establish the almost sure convergence rates of our estimators. When the underlying outcome regression function is a piecewise function of the treatment, we further derive the almost sure convergence rates of the estimated change point locations, and show that with probability 1, the number of change points can be correctly estimated with sufficiently large sample size. These findings are nontrivial extensions of classical results derived for models without covariates. For instance, deriving the asymptotic behavior of change point estimators for these models typically relies on the tail inequalities for the partial sum process (see e.g., Frick et al. 2014). However, these technical tools are not directly applicable to our settings where deep learning is adopted to model the outcome regression function. Moreover, we expect our theories to also be of general interest to the line of work on developing theories for deep learning methods (see e.g., Imaizumi and Fukumizu 2019; Schmidt-Hieber et al. 2020; Farrell et al. 2021).

The rest of this chapter is organized as follows. In Section 2.2, we introduce the statistical framework, define the notion of I2DR, and posit our working model assumptions. In Section 2.3, we propose the jump interval-learning method and discuss its detailed implementation. Statistical properties of the proposed I2DR and the estimator for the mean outcome under the proposed I2DR are presented in Section 2.4. We further develop a confidence interval for the expected outcome under the estimated I2DR. Simulation studies are conducted in Section 2.5 to evaluate the finite sample performance of our proposed method. We apply our method to a real dataset from a warfarin study in Section 2.6. All the proofs and additional discussions are provided in Appendix A.
2.2 Statistical Framework

This section is organized as follows. We first introduce the model setup in Section 2.2.1. The definition of I2DR is formally presented in Section 2.2.2. In Section 2.2.3, we posit two working models assumptions for the expected outcome as a function the treatment level. We aim to develop a method that works under both working assumptions.

2.2.1 Model Setup

We begin with some notations. Let $A$ denote the treatment level assigned to a randomly selected individual in the population from a compact interval. Without loss of generality, suppose $A$ belongs to $[0, 1]$. Let $X \in \mathbb{X}$ be that individual's baseline covariates where the support $\mathbb{X}$ is a subset in $\mathbb{R}^p$. We assume the covariance matrix of $X$ is positive definite. Let $Y \in \mathbb{R}$ denote that individual's associated outcome, the larger the better by convention. Let $p(\cdot | x)$ denote the probability density function of $A$ given $X = x$. In addition, for any $a \in [0, 1]$, define the potential outcome $Y^*(a)$ as the outcome of that individual that would have been observed if they were receiving treatment $a$. The observed data consists of the covariate-treatment-outcome triplets $\{(X_i, A_i, Y_i) : i = 1, \ldots, n\}$ where $(X_i, A_i, Y_i)$'s are i.i.d. copies of $(X, A, Y)$. Based on this data, we wish to learn an optimal decision rule to possibly maximize the expected outcome of future subjects using their baseline information.

Formally speaking, an individualized decision rule (IDR) is a deterministic function $d(\cdot)$ that maps the covariate space $\mathbb{X}$ to the treatment space $[0, 1]$. The optimal IDR is defined to maximize the expected outcome (value function) $V(d) = \mathbb{E}\{Y^*(d(X))\}$ among all IDRs. The following assumptions guarantee the optimal IDR is identifiable from the observed data.

(A1.) Stable Unit Treatment Value Assumption (SUTVA): $Y = Y^*(A)$, almost surely,
(A2.) No unmeasured confounders: $\{Y^*(a) : a \in [0, 1]\} \perp \perp A | X$,
(A3.) Positivity: there exists some constant $c_* > 0$ such that $p(a|x) \geq c_*$ for any $x \in \mathbb{X}$ and $a \in [0, 1]$.

SUTVA requires that the outcome of each individual depends on their own treatment only. In other words, there is no interference effect between individuals. Assumption (A2) requires that the baseline covariates have included enough confounders given which the potential outcomes and the received treatment are independent. (A2) and (A3) automatically hold in randomized studies. These three assumptions are commonly imposed in the literature for estimating an optimal IDR (see e.g., Chen et al. 2016; Zhu et al. 2020a; Schulz and Moodie 2020). Under (A1)-(A3), we have $V(d) = \mathbb{E}\{Q(X, d(X))\}$ where $Q(x, a) = \mathbb{E}(Y|X = x, a)$. 


\( x, A = a \) is the conditional mean of an individual’s outcome given their received treatment and baseline covariates. We refer to this function as the outcome regression function. As a result, the optimal IDR for an individual with covariates \( x \) is given by \( \text{arg max}_{a \in [0,1]} Q(x, a) \). Let \( V^{opt} \) denote the value function under the optimal IDR. We have \( V^{opt} = E\{\sup_{a \in [0,1]} Q(X, a)\} \).

### 2.2.2 I2DR

The focus of this chapter is to develop an optimal individualized interval-based decision rule (I2DR). As commented in the introduction, these decision rules are more flexible to implement in practice when compared to single-valued decision rules in personalized dose finding and dynamic pricing.

We define an I2DR as a function \( d(\cdot) \) that takes an individual’s covariates \( x \) as input and outputs an interval \( \mathcal{I} \subseteq [0,1] \). Given the recommended interval \( \mathcal{I} \), different doctors / agents might assign different treatments to patients / products. The actual treatments that subjects receive in the population will have a distribution function \( \Pi^*(\cdot; x, \mathcal{I}) \). Throughout this chapter, we assume \( \Pi^*(\cdot; x, \mathcal{I}) \) has a bounded density function \( \pi^*(\cdot; x, \mathcal{I}) \) for any \( x \) and \( \mathcal{I} \). Apparently, we have \( \int_{\mathcal{I}} \pi^*(a; x, \mathcal{I}) da = 1 \), for any interval \( \mathcal{I} \) and \( x \in X \). When (A1)-(A3) hold, the associated value function under an I2DR \( d(\cdot) \) equals

\[
V^*(d) = E\left( \int_{d(X)} Q(X, a) \pi^*(a; X, d(X)) da \right).
\]

Restricting \( d(\cdot) \) to be a scalar-valued function, \( V^*(d) \) is reduced to \( V(d) \).

Given the dataset, one may estimate \( V^*(d) \) nonparametrically for any \( d(\cdot) \) and directly search the optimal I2DR based on the estimated value function. However, such a value search method has the following two limitations. First, a nonparametric estimator of \( V^*(d) \) requires to specify the preference function \( \pi^* \), which might be unknown to us. Second, even though a nonparametric estimator of \( V^*(d) \) can be derived, it remains unknown how to efficiently compute the I2DR that maximizes the estimated value (see Section 2.7.2 for details). To overcome these limitations, we propose a semiparametric model for the outcome regression function and use a model-assisted approach to derive the optimal I2DR. We formally introduce our method in Section 2.3.
2.2.3 Working Model Assumptions

In this section, we introduce two working models for the outcome regression function, corresponding to a piecewise function and a continuous function of the treatment level.

**Model I (Piecewise Functions).** Suppose

\[
Q(x, a) = \sum_{\mathcal{I} \in \mathcal{P}_0} q_{\mathcal{I},0}(x) I(a \in \mathcal{I}) \quad \forall x \in X, a \in [0, 1],
\]

for some partition \( \mathcal{P}_0 \) of \([0, 1]\) and a collection of continuous functions \((q_{\mathcal{I},0})_{\mathcal{I} \in \mathcal{P}_0}\), where the number of intervals in \( \mathcal{P}_0 \) is finite. Specifically, a partition \( \mathcal{P} \) of \([0, 1]\) is defined as a collection of mutually disjoint intervals \(\{[\tau_0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{K-1}, \tau_K]\}\) for some \(0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \tau_K = 1\) and some integer \(K \geq 1\). As commented in our introduction, we expect the above model assumption holds in real-world examples such as dynamic pricing.

**Model II (Continuous Functions).** Suppose \(Q(x, a)\) is a continuous function of \(a\) and \(x\), for any \(x \in X\) and \(a \in [0, 1]\).

We aim to propose an optimal I2DR that optimizes the value function when either Model I or Model II holds.

2.3 Methods

In this section, we first present the proposed jump interval-learning and its motivation in Section 2.3.1. We next introduce two concrete proposals, i.e., linear jump interval-learning and deep jump interval-learning, to detail our methods in Section 2.3.2. We then present the dynamic programming algorithm to implement jump interval-learning (see Algorithm 1 for an overview) in Section 2.3.3. Finally, we provide more details on tuning parameter selection in Section 2.3.4.

2.3.1 Jump Interval-learning

We use Model I to present the motivation of our jump interval-learning. In view of (2.1), any treatment level within an interval \(\mathcal{I} \in \mathcal{P}_0\) will yield the same efficacy to a given individual. The optimal I2DR is then given by

\[
d^{opt}(x) = \arg\max_{\mathcal{I} \in \mathcal{P}_0} q_{\mathcal{I},0}(x),
\]
independent of the preference function $\pi^*$. To see this, notice that

$$V^{\pi^*}(d^{\text{opt}}) = \mathbb{E} \left( \int_{d^{\text{opt}}(X)} \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(a \in \mathcal{I}) \pi^*(a; X, d^{\text{opt}}(X)) da \right)$$

$$= \mathbb{E} \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(d^{\text{opt}}(X) \in \mathcal{I}) \int_{d^{\text{opt}}(X)} \pi^*(a; X, d^{\text{opt}}(X)) da.$$ 

For any I2DR $d(\cdot)$, we have $\int_{d^{\text{opt}}(X)} \pi^*(a; X, d^{\text{opt}}(X)) da = \int_{d(X)} \pi^*(a; X, d(X)) da = 1$ by definition. It follows that

$$V^{\pi^*}(d^{\text{opt}}) = \mathbb{E} \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(d^{\text{opt}}(X) \in \mathcal{I}) \int_{d(X)} \pi^*(a; X, d(X)) da$$

$$\geq \mathbb{E} \int_{d(X)} \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(a \in \mathcal{I}) \pi^*(a; X, d(X)) da = V^{\pi^*}(d),$$

where the inequality is due to that

$$Q(X, d^{\text{opt}}(X)) = \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(d^{\text{opt}}(X) \in \mathcal{I}) \geq \sum_{\mathcal{I} \in \mathcal{P}} q_{\mathcal{I}, 0}(X) \mathbb{I}(a \in \mathcal{I}) = Q(X, a),$$

almost surely for any $a \in [0, 1]$. Therefore, to derive the optimal I2DR, it suffices to estimate $q_{\mathcal{I}, 0}(\cdot)$. For notation simplicity, in the rest of this chapter, we denote $V^{\pi^*}(d)$ by $V(d)$ for any decision rule $d$.

From now on, we focus on a subset of intervals in $[0, 1]$. By *interval* we always refer to those of the form $[a, b]$ for some $0 \leq a < b < 1$ or $[a, 1]$ for some $0 \leq a < 1$. For any partition $\mathcal{P} = \{[0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{K-1}, 1]\}$, we use $J(\mathcal{P})$ to denote the set of change point locations, i.e., $\{\tau_1, \tau_2, \ldots, \tau_{K-1}\}$, and $|\mathcal{P}|$ to denote the number of intervals in $\mathcal{P}$. Our proposed method yields a partition $\widehat{\mathcal{P}}$ and an I2DR $\widehat{d}(\cdot)$ such that $\widehat{d}(x) \in \widehat{\mathcal{P}}, \forall x \in X$. The number of intervals in $\widehat{\mathcal{P}}$ (denoted by $|\widehat{\mathcal{P}}|$) involves a trade-off. If $|\widehat{\mathcal{P}}|$ is too large, then $\widehat{\mathcal{P}}$ will contain many short intervals, making the resulting decision rule hard to implement in practice. Yet, a smaller value of $|\widehat{\mathcal{P}}|$ might result in a smaller value function. Our proposed method adaptively determines $|\widehat{\mathcal{P}}|$ based on jump-penalized regression.

We next detail our method. Jump interval-learning consists of the following two steps. In the first step, we estimate the outcome regression function using jump penalized least squares regression. Then we derive the corresponding I2DR from the resulting estimator.
To begin with, we cut the entire treatment range into $m$ initial intervals:

$$[0, 1/m), [1/m, 2/m), \ldots, [(m-1)/m, 1]. \quad (2.2)$$

The integer $m$ is allowed to diverge with the number of observations $n$. For instance, it can be specified by the clinical physician such that the output dose interval for each individual is at least of the length $m^{-1}$. When no prior knowledge is available, we recommend to set $m$ to be proportional to $n$. It is worth mentioning that (2.2) is not the final partition that we recommend. Nor is it equal to $\mathcal{P}_0$ defined in Model I. Given (2.2), we are looking for a partition $\widetilde{\mathcal{P}}$ such that each interval in $\widetilde{\mathcal{P}}$ corresponds to a union of some of the these $m$ intervals. In other words, we will adaptively combine some of these intervals to form $\widetilde{\mathcal{P}}$.

More specifically, let $\mathcal{B}(m)$ denote the set of partitions $\mathcal{P}$ that satisfy the following requirement: the end-points of each interval $\mathcal{I} \in \mathcal{P}$ lie on the grid $\{j/m : j = 0, 1, \ldots, m\}$. We associate to each partition $\mathcal{P} \in \mathcal{B}(m)$ a collection of functions $\{q_\mathcal{I}(.; \theta_\mathcal{I})\}_{\mathcal{I} \in \mathcal{P}}$ for $\mathcal{Q}$ as some class of functions, where $\theta_\mathcal{I}$ is the underlying parameter associated to interval $\mathcal{I}$. We propose to estimate $\widetilde{\mathcal{P}}$ by solving

$$\left(\widetilde{\mathcal{P}}, \{q_\mathcal{I} : \mathcal{I} \in \widetilde{\mathcal{P}}\}\right) = \arg\min_{\mathcal{P} \in \mathcal{B}(m)} \left\{ \sum_{\mathcal{I} \in \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \left( Y_i - q_\mathcal{I}(X_i; \theta_\mathcal{I}) \right)^2 + \lambda_n |\mathcal{I}| \|\theta_\mathcal{I}\|_2^2 + \gamma_n |\mathcal{P}| \right\}, \quad (2.3)$$

where $\lambda_n$ and $\gamma_n$ are some nonnegative regularization parameters specified in Section 2.3.4, and $\|\theta_\mathcal{I}\|_2^2$ denote the Euclidean norm of the model parameter $\theta_\mathcal{I}$. The purpose of introducing the $\ell_2$-type penalty term $\lambda_n |\mathcal{I}| \|\theta_\mathcal{I}\|_2^2$ is to help to prevent overfitting in large $p$ problems. The purpose of introducing the $\ell_0$-type penalty term $\gamma_n |\mathcal{P}|$ is to control the total number of jumps. When $m = n$, $\lambda_n = 0$, $A_i = i/n$, $\forall 1 \leq i \leq n$, no baseline covariates are collected, the above optimization corresponds to the jump-penalized least square estimator proposed by Boysen et al. (2009). We refer to this step as jump interval-learning (JIL).

For a fixed $\mathcal{P}$, solving the optimization function in (2.3) yields its associated outcome regression functions $\{q_\mathcal{I}\}_{\mathcal{I} \in \mathcal{P}}$. This step involves parametric or nonparametric regression and can be solved via existing statistical or machine learning approaches. We provide two concrete study cases below, based on linear regression and deep learning. These estimated outcome regression functions can be viewed as functions of $\mathcal{P}$. As such, $\widetilde{\mathcal{P}}$ is adaptively determined by minimizing the penalized least square function in (2.3).
To maximize the expected outcome of interest, our proposed I2DR is then given by

\[
\hat{d}(x) = \arg\max_{\mathcal{I} \in \mathcal{P}} \tilde{q}_\mathcal{I}(x), \quad \forall x \in \mathbf{X}.
\]  

(2.4)

When the argmax in (2.4) is not unique, \(\hat{d}(\cdot)\) outputs the interval that contains the smallest treatment.

We next evaluate the value function under the proposed I2DR \(V(\hat{d})\) and \(V(d^{opt})\). For each interval \(\mathcal{I}\) in the estimated optimal partition \(\mathcal{P}\), we estimate the generalized propensity score function \(e(\mathcal{I}|x) \equiv \Pr(A \in \mathcal{I}|X = x)\). Let \(\hat{e}(\mathcal{I}|x)\) denote the resulting estimate. Following the estimation strategy in Zhang et al. (2012a), we propose the following value estimator under (2.4),

\[
\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\mathbb{I}\{A_i \in \hat{d}(X_i)\}}{\hat{e}(\hat{d}(X_i)|X_i)} \{ Y_i - \max_{\mathcal{I} \in \mathcal{P}} \tilde{q}_\mathcal{I}(X_i) \} + \max_{\mathcal{I} \in \mathcal{P}} \tilde{q}_\mathcal{I}(X_i) \right].
\]  

(2.5)

Statistical properties of the estimates in (2.4) and (2.5) are studied in Section 2.4. Although we use the example of piecewise functions to motivate our procedure, the proposed method allows the outcome regression function to be a continuous function of \(a\) and \(x\) as well. See Section 2.4 for detail.

### 2.3.2 Linear- and Deep-JIL

In practice, we consider two concrete proposals to implementing jump interval-learning, by considering a linear function class and a deep neural networks class for \(\mathcal{Q}_\mathcal{I}\) in (2.3).

**Case 1: Linear-JIL**

We use a linear regression model for \(\mathcal{Q}_\mathcal{I}\). Specifically, we set \(q(x, \theta_\mathcal{I})\) to \(\tilde{x}^\top \theta_\mathcal{I}\) for any interval \(\mathcal{I}\) and \(x \in \mathbf{X}\), where \(\tilde{x}\) is a shorthand for the vector \((1, x^\top)^\top\). Adopting the linearity assumption, we have \(\tilde{q}_\mathcal{I}(x) = \tilde{x}^\top \hat{\theta}_\mathcal{I}\) for some \(\hat{\theta}_\mathcal{I}\). It follows from (2.4) that the proposed I2DR corresponds to a linear decision rule, i.e., \(\hat{d}(x) = \arg\max_{\mathcal{I} \in \mathcal{P}} \tilde{x}^\top \hat{\theta}_\mathcal{I}\). As such, the linearity assumption ensures our I2DR is interpretable to the domain experts.

We next discuss how to compute \(\mathcal{P}\) and \(\{\hat{\theta}_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}\). The objective function in (2.3) is
reduced to

\[
(\tilde{\mathcal{T}}, \{\tilde{\theta}_I : I \in \tilde{\mathcal{T}}\}) = \\
= \arg\min_{(\mathcal{P} \in \mathcal{B}(m), \{\theta_I : I \in \mathcal{P}\})} \left\{ \sum_{I \in \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in I)(Y_i - \bar{X}_I^\top \theta_I)^2 + \lambda_n |I||\theta_I|^2_2 + \gamma_n |\mathcal{P}| \right\},
\]

where \( \bar{X}_I = (1, X_i^\top)^\top \). We refer to this step as linear jump interval-learning (L-JIL). The ridge penalty \( \lambda_n |\mathcal{I}||\theta_I|^2_2 \) in (2.6) guarantees that for any interval \( I \in \tilde{\mathcal{T}} \), the parameter \( \tilde{\theta}_I \) is well defined even when \( \sum_{i=1}^{n} \mathbb{I}(A_i \in I) < p + 1 \) such that the matrix \( \sum_{i=1}^{n} \mathbb{I}(A_i \in I) \bar{X}_I \bar{X}_I^\top \) is not invertible. It also prevents over-fitting and yields more accurate estimate in high-dimensional settings.

**Case 2: Deep-JIL**

We next consider using deep neural networks (DNNs) to approximate the outcome regression function, so as to capture the complex dependence between the outcome and covariates. Specifically, the network consists of \( p \) input units (coloured in blue in Figure 2.1), corresponding to the covariates \( X \). The hidden units (coloured in green) are grouped in a sequence of \( L \) layers. Each unit in the hidden layer is determined as a nonlinear transformation of a linear combination of the nodes from the previous layer. The total number of parameters in the network is denoted by \( W \). See Figure 2.1 for an illustration.

The parameters in DNNs can be solved using a stochastic gradient descent algorithm. In our implementation, we apply the Multi-layer Perceptron (MLP) regressor Pedregosa et al. (2011) for parameter estimation. We refer to the resulting optimization as deep jump interval-learning (D-JIL).

Finally, we remark that alternative to our approach, one may directly apply DNN that takes the covariate-treatment pair \((X, A)\) as the input to learn the outcome regression function. However, the resulting estimator for the outcome regression function is not guaranteed to be a piecewise function of the treatment. As such, it cannot yield an I2DR.

### 2.3.3 Implementation

In this section, we present the computational details for jump interval-learning. We employ the dynamic programming algorithm (see e.g., Friedrich et al. 2008) to find the optimal partition \( \tilde{\mathcal{T}} \) that minimizes the objective function (2.3). Meanwhile, other algorithms for multi-scale change point detection are equally applicable (see e.g., Scott and Knott
Figure 2.1: Illustration of DNN with $L = 2$ and $W = 25$; here $\mu \in \mathbb{R}^p$ is the input, the output is given by $A^{(3)}\sigma(A^{(2)}\sigma(A^{(1)}\mu + b^{(1)}) + b^{(2)}) + b^{(3)}$ where $A^{(l)}$, $b^{(l)}$ denote the corresponding parameters to produce the linear transformation for the $(l-1)$th layer and that $\sigma$ denotes the componentwise ReLU function. In this example, $W = \sum_{j=1}^{3}(\|A^{(3)}\|_0 + \|b^{(3)}\|_0) = 25$ where $\| \cdot \|_0$ denotes the number of nonzero elements in the vector or matrix.

1974; Harchaoui and Lévy-Leduc 2010; Fryzlewicz 2014). Specifically, we adopt the PELT method proposed by Killick et al. (2012a) that includes additional pruning steps within the dynamic programming framework to achieve a linear computational cost. Given $\mathcal{P}$, the set of functions $\{\hat{q}_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}$ can be computed via either linear regression or deep neural network.

To detail our procedure, for any interval $\mathcal{I} \in [0, 1]$, we define the cost function

$$\text{cost}(\mathcal{I}) = \min_{q(\cdot; \theta_\mathcal{I}) \in \mathcal{Q}_\mathcal{I}} \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \{Y_i - q(X_i; \theta_\mathcal{I})\}^2 + \lambda_n \|\theta_\mathcal{I}\|_2^2 \right],$$

where $\mathcal{Q}_\mathcal{I}$ is a class of linear functions or deep neural networks, corresponding to L-JIL and D-JIL, respectively.

For any integer $1 \leq r < m$, denote by $\mathcal{B}(m, r)$ the set consisting of all possible partitions $\mathcal{P}_r$ of $[0, r/m)$ such that the end-points of each interval $\mathcal{I} \in \mathcal{P}_r$ lie on the grid $\{j/m : j = 0, 1, \ldots, r\}$. Set $\mathcal{B}(m, m) = \mathcal{B}(m)$, we define the Bellman function

$$B(r) = \inf_{\mathcal{P}_r \in \mathcal{B}(m, r)} \left( \sum_{\mathcal{I} \in \mathcal{P}_r} \text{cost}(\mathcal{I}) + \gamma_n (|\mathcal{P}_r| - 1) \right).$$

Let $B(0) = -\gamma_n$, the dynamic programming algorithm relies on the following recursion
Global: data \{(X_i, A_i, Y_i): i = 1, \ldots, n\}; sample size \(n\); covariates dimension \(p\); number of initial intervals \(m\); penalty terms \(\gamma_n\).

Local: integers \(l, r \in \mathbb{N}\); cost dictionary \(\mathcal{C}\); a vector of integers \(\tau \in \mathbb{N}^m\); Bellman function \(B \in \mathbb{R}^m\); a set of candidate point lists \(\mathcal{R}\).

Output: \(\mathcal{P}\) and \(\{\hat{q}_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}\).

I. Initialization.
1. Set \(B(0) \leftarrow -\gamma_n\); \(\mathcal{P} \leftarrow \text{null}\); \(\tau \leftarrow \text{null}\); \(\mathcal{R}(0) \leftarrow \{0\}\).
2. Define the cost function \(\mathcal{C}(\mathcal{I})\):
   (i). If \(\mathcal{C}(\mathcal{I}) \leftarrow \text{null}\):
      (a). Apply Linear / MLP regression: \(\hat{q}_\mathcal{I}(\cdot) \leftarrow \arg \min_q \sum_i \mathbb{I}(A_i \in \mathcal{I}) \{Y_i - q(X_i)\}^2\);
      (b). Calculate the cost: \(\mathcal{C}(\mathcal{I}) \leftarrow \sum_i \mathbb{I}(A_i \in \mathcal{I}) \{\hat{q}_\mathcal{I}(X_i) - Y_i\}^2\);
   (ii). Return \(\mathcal{C}(\mathcal{I})\).

II. Apply the PELT method. For \(r = 1, \ldots, m\):
1. \(B(r) = \min_{j \in \mathcal{R}(r)} \{B(j) + \mathcal{C}([j/m, r/m]) + \gamma_n\}\);
2. \(j^* \leftarrow \arg \min_{j \in \mathcal{R}(r)} \{B(j) + \mathcal{C}([j/m, r/m]) + \gamma_n\}\);
3. \(\tau(r) \leftarrow \{j^*, \tau(j^*)\}\);
4. \(\mathcal{R}(r) \leftarrow \{j \in \mathcal{R}(r-1) \cup \{r-1\} : B(j) + \mathcal{C}([j/m, (r-1)/m]) \leq B(r-1)\}\);

III. Get Partitions. \(\tau^* \leftarrow \tau(m); r \leftarrow m\); \(l \leftarrow \tau^*[r]\); While \(r > 0\):
1. Let \(\mathcal{I} = [l/m, r/m]\) if \(r < m\) else \(\mathcal{I} = [l/m, 1]\);
2. \(\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{I}\);
3. \(\hat{q}_\mathcal{I}(\cdot) \leftarrow \arg \min_{q} \sum_i \mathbb{I}(A_i \in \mathcal{I}) \{Y_i - q(X_i)\}^2\);
4. \(r \leftarrow l; l \leftarrow \tau^*[r]\);

return \(\mathcal{P}\) and \(\{\hat{q}_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}\).

Algorith 1: Jump interval-learning

\[ B(r) = \min_{j \in \mathcal{R}_r} \{B(j) + \gamma_n + \text{cost}([j/m, r/m])\}, \quad \forall r \geq 1. \]  (2.7)

where \(\mathcal{R}_r\) is the candidate change points list updated by

\[ \{j \in \mathcal{R}_{r-1} \cup \{r-1\} : B(j) + \text{cost}([j/m, (r-1)/m]) \leq B(r-1)\}, \]  (2.8)

during each iteration with \(\mathcal{R}_0 = \{0\}\). The constraint listed in (2.8) iteratively updates the set of candidate change points and removes values that can never be the minima of the objective function. It speeds up the computation, leading to a cost that is linear in the number of observations (Killick et al. 2012a).
We briefly summarize our algorithm below. For a given integer \( r \), we search the optimal change point location \( j \) that minimizes the above Bellman function \( B(r) \) in (2.7). This requires to apply the linear / MLP regression to learn \( \hat{q}_{j/m, r/m} \) and \( \text{cost}(\{j/m, r/m\}) \) for each \( j \in R \). Let \( j^* \) be the corresponding minimizer. We then define the change points list \( \tau(r) = \{j^*, \tau(j^*)\} \). This procedure is iterated to compute \( B(r) \) and \( \tau(r) \) for \( r = 1, \ldots, m \). The optimal partition \( \mathcal{P} \) is determined by the values stored in \( \tau(\cdot) \). A pseudocode containing more details is given in Algorithm 1.

### 2.3.4 Tuning Parameters

Our proposal requires to specify the tuning parameters \( m, \lambda_n \) and \( \gamma_n \). We first discuss the choice of \( m \). In practice, we recommend to set \( m = n/c \) with some constant \( c > 0 \) such that \( m \) and \( n \) are of the same order. In our simulation studies, we tried several different values of \( c \) and found the resulting estimated I2DRs have approximately the same value function. Thus, the proposed I2DR is not overly sensitive to the choice of this constant. Detailed empirical results can be found in Section 2.5.3.

We next discuss the choices of \( \lambda_n \) and \( \gamma_n \). Selection of these tuning parameters relies on the concrete proposal to approximate the outcome regression function. We elaborate below.

**Tuning in L-JIL**

For L-JIL, we choose \( \gamma_n \) and \( \lambda_n \) simultaneously via cross-validation. The theoretical requirements of \( \gamma_n \) and \( \lambda_n \) for L-JIL are imposed in the statement of Theorem 1. We further develop an algorithm that substantially reduces the computation complexity resulting from the use of cross-validation.

To be more specific, let \( \Lambda_n = \{\lambda_n^{(1)}, \cdots, \lambda_n^{(H)}\} \) and \( \Gamma_n = \{\gamma_n^{(1)}, \cdots, \gamma_n^{(J)}\} \) be the set of candidate tuning parameters. For a given integer \( K_0 \), we randomly split the data into \( K_0 \) equal sized subgroups. Let \( G_k \) denote indices of the subsamples in the \( k \)th subgroup, for \( k = 1, \cdots, K_0 \). Let \( G_{-k} \) denote the complement of \( G_k \). For any \( \lambda_n \in \Lambda_n, \gamma_n \in \Gamma_n, k \in \{1, \cdots, K_0\}, \) let \( (\mathcal{P}_{\lambda_n, \gamma_n, k}, \{\tilde{\theta}_{\mathcal{S}, \gamma_n, \lambda_n, k} : \mathcal{S} \in \mathcal{P}_{\lambda_n, \gamma_n, k}\}) \) denote the optimizer (2.6), computed based on the data in \( G_{-k} \). We aim to choose \( \gamma_n \) and \( \lambda_n \) that minimizes

\[
\frac{1}{n} \sum_{k=1}^{K_0} \sum_{i \in G_k} \sum_{\mathcal{S} \in \mathcal{P}_{\lambda_n, \gamma_n, k}} \mathbb{I}(A_i \in \mathcal{S}) \{Y_i - \overline{X_i}^\top \tilde{\theta}_{\mathcal{S}, \gamma_n, \lambda_n, k}\}^2.
\]  

(2.9)
To solve (2.9), we remark that there is no need to apply Algorithm 1 \(|\Lambda_n| \times |\Gamma_n|\) times to compute the minimizer of (2.6) over the set of candidate tuning parameters. We develop an algorithm to facilitate the computation. The key observation is that, for any interval \(\mathcal{I} \subseteq [0, 1]\) and \(k \in \{1, \cdots, K_0\}\), the set of estimators \(\{\hat{\theta}_{\mathcal{I}, \gamma_n, \lambda_n, k} : \gamma_n \in \Gamma_n, \lambda_n \in \Lambda_n\}\) can be obtained simultaneously over the set of candidate tuning parameters. This forms the basis of our algorithm. More details are provided in Section A.1 of the supplementary article.

**Tuning in D-JIL**

As for D-JIL, we find that the MLP regressor is not overly sensitive to the choice of \(\lambda_n\), so we set \(\lambda_n = 0\). The parameter \(\gamma_n\) is chosen based on cross-validation. The theoretical requirement of \(\gamma_n\) for D-JIL is imposed in the statement of Theorem 2. To implement the cross-validation, we randomly split the data into \(K_0\) equal sized subgroups, denoted by \(\{(X_i, A_i, Y_i)\}_{i \in G_1}, \{(X_i, A_i, Y_i)\}_{i \in G_2}, \cdots, \{(X_i, A_i, Y_i)\}_{i \in G_{K_0}}\), accordingly. For each \(\gamma_n\) and \(k = 1, \cdots, K_0\), we compute the estimators \(\hat{\theta}_{\mathcal{I}, \gamma_n, k}\) and \(\hat{q}_{\mathcal{I}, \gamma_n, k}(\cdot)\) based on the sub-dataset in \(G_k\). Then we choose \(\gamma_n\) that minimizes

\[
\frac{1}{n} \sum_{k=1}^{K_0} \sum_{i \in G_k} \sum_{\mathcal{I} \in \mathcal{I}_{\gamma_n, k}} \mathbb{I}(A_i \in \mathcal{I}) \{Y_i - \hat{q}_{\mathcal{I}, \gamma_n, k}(X_i)\}^2.
\]

We also remark that other tuning parameters, such as the learning rate, the numbers of hidden nodes and hidden layers, are set to the default values of the MLP regressor implementation (Pedregosa et al. 2011).

**2.4 Theory**

We establish the statistical properties of our proposed method in this section. As we have commented, we consider both cases where the outcome regression function is either a piecewise or continuous function of the treatment. We first study the statistical properties of L-JIL and D-JIL under Model I, respectively. We next outline a procedure to construct a confidence interval for the value under the proposed I2DR and prove its validity under these two methods. Finally, we investigate the properties of our proposed method under Model II. The theoretical results justify that our method will work when the outcome regression function is either piecewise or continuous function.
2.4.1 Properties under Model I

Results for L-JIL

To establish the theoretical properties of the I2DR obtained by L-JIL, we first assume (2.1) holds with \( q_{I_0,0}(x) = \bar{x}^\top \theta_{I_0,0} \) for any \( x \in X \) and \( I \in \mathcal{P}_0 \). In other words, the outcome regression function \( Q(x, a) \) is linear in \( x \) and piecewise constant in \( a \). Without loss of generality, assume \( \theta_{0,I_1} \neq \theta_{0,I_2} \) for any two adjacent intervals \( I_1, I_2 \in \mathcal{P}_0 \). This guarantees that the representation in (2.1) is unique. We write \( a_n \approx b_n \) for two sequences \( \{a_n\}, \{b_n\} \) if there exists some universal constant \( c \geq 1 \) such that \( c^{-1}b_n \leq a_n \leq cb_n \). Define \( \theta_0(\cdot) = \sum_{I \in \mathcal{P}_0} \theta_{I_0,0}(\cdot | I) \).

Giving \((c_P, \{b_{I_0} : I \in c_P\})\), our estimator for the function \( \theta_0(\cdot) \) is defined by

\[
\hat{\theta}(\cdot) = \sum_{I \in \widehat{\mathcal{P}}} \hat{\theta}_I(\cdot | I).
\] (2.10)

This yields a piecewise constant approximation of \( \theta_0(\cdot) \). We first study the theoretical properties of \( \hat{\theta}(\cdot) \). Toward that end, we need to impose the following condition on the probability tails of \( X \) and \( Y \).

(A4) Suppose there exists some constant \( \omega > 0 \) such that

\[
\|X^{(j)}\|_{\psi_2|A} \leq \omega, \text{ for any } j \in \{1, \ldots, p\}, \text{ almost surely and } \|Y\|_{\psi_2|A} \leq \omega \text{ almost surely,}
\]

where \( X^{(j)} \) denotes the \( j \)th element of \( X \), and that for any random variable \( Z \), \( \|Z\|_{\psi_2|A} \) denotes the conditional Orlicz norm given the treatment \( A \),

\[
\|Z\|_{\psi_2|A} \triangleq \inf_{C > 0} \mathbb{E}\left\{ \exp\left( \frac{|Z|^2}{C^2} \right) \bigg| A \right\} \leq 2.
\]

We remark that Condition (A4) is automatically satisfied when the covariates and the outcomes are bounded.

**Theorem 1** Assume (A1)-(A4) hold and (2.1) holds with \( q_{\mathcal{P},0}(x) = \bar{x}^\top \theta_{\mathcal{P},0} \). Assume \( A \) has a bounded probability density function on \([0, 1]\). Assume \( m \approx n, \lambda_n = O(n^{-1} \log n) \), \( \{\gamma_n\}_{n \in \mathbb{N}} \) satisfies \( \gamma_n \to 0 \) and \( \gamma_n n / \log n \to \infty \). Then, there exists some constant \( \hat{c} > 0 \) such that the following events hold with probability at least \( 1 - O(n^{-2}) \):

(i) \( |\widehat{\mathcal{P}}| = |\mathcal{P}_0| \).

(ii) \( \max_{\tau \in (\mathcal{P}_0)} \min_{\hat{\tau} \in (\widehat{\mathcal{P}})} |\hat{\tau} - \tau| \leq \hat{c} n^{-1} \log n. \)

(iii) \( \int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|^2_2 \, da \leq \hat{c} n^{-1} \log n. \)

In Theorem 1, results in (i) show the model selection consistency of our jump penalized estimator. Results in (ii) imply that the estimated change point locations converge at a rate
of $O_p(n^{-1} \log n)$. In (iii), we derive an upper error bound for the integrated $\ell_2$ loss of $\hat{\theta}(\cdot)$. As discussed in the introduction, derivation of Theorem 1 is nontrivial. A number of technical lemmas (see Lemma 1-4 in Section A.2.1) are established to prove Theorem 1. These results can be easily extended to study general covariate-based change point models.

We next establish the convergence rate of $V_{opt} - V(d)$, where $V_{opt} = V(d_{opt})$. The quantity $V_{opt} - V(d)$ represents the difference between the optimal value and the value under the proposed I2DR. The smaller the difference, the better the I2DR. Notice that $V_{opt} \geq V(d)$ for any I2DR $d(\cdot)$. It suffices to provide an upper bound for $V_{opt} - V(d)$. We impose the following condition.

(A5.) Assume for any $\mathcal{I}_1, \mathcal{I}_2 \in \mathcal{P}_0$, there exist some constants $\gamma, \delta_0 > 0$ such that

$$\Pr(0 < |q_{\mathcal{I}_1,0}(X) - q_{\mathcal{I}_2,0}(X)| \leq t) = O(t^\gamma),$$

where the big-$O$ term is uniform in $0 < t \leq \delta_0$.

Condition (A5) is commonly assumed in the literature to derive sharp convergence rate for the value function under the estimated optimal IDR (Qian and Murphy 2011; Luedtke and Van Der Laan 2016; Shi et al. 2020). It is very similar to the margin condition (Tsybakov 2004; Audibert and Tsybakov 2007) used in the classification literature. This condition is automatically satisfied with $\gamma = 1$ when $q_{\mathcal{I},0}(X)$ has a bounded probability density function for any $\mathcal{I} \in \mathcal{P}_0$.

**Theorem 2** Assume the conditions in Theorem 1 are satisfied. Further assume (A5) holds. Then, we have

$$V_{opt} - V(d) \leq \bar{c} (n^{-1} \log n)^{(1+\gamma)/2} + \bar{c} n^{-1} \log n,$$

for some constant $\bar{c} > 0$, with probability at least $1 - O(n^{-2})$.

When (A5) holds with $\gamma = 1$, Theorem 2 suggests that $V(d)$ converges to the optimal value at a rate of $O_p(n^{-1})$ up to some logarithmic factor. Notice that the events defined in Theorem 1 and 2 occur with probability at least $1 - O(n^{-2})$. Since $\sum_{n \geq 1} n^{-2} < +\infty$, an application of Borel-Cantelli lemma implies that these events will occur for sufficiently large $n$ almost surely.
Results for D-JIL

We study the theoretical properties of the proposed I2DR under Model I when D-JIL is applied. Similar to the linear case, we assume \( q_{\mathcal{I}_1,0} \neq q_{\mathcal{I}_2,0} \) for any two adjacent intervals \( \mathcal{I}_1, \mathcal{I}_2 \in \mathcal{P}_0 \). For any \( \mathcal{I} \), we set the regression class \( \mathcal{Q}_\mathcal{I} \) to a class of DNN with \( L_\mathcal{I} \) hidden layers and \( W_\mathcal{I} \) many number of parameters.

To derive the theoretical properties of D-JIL, we assume the outcome regression function is a smooth function of the baseline covariates (see assumption (A6) below). Meanwhile, D-JIL is valid when \( Q(\cdot, a) \) is a nonsmooth function of \( x \) as well (see e.g., Imaizumi and Fukumizu 2019). Specifically, define the class of \( \beta \)-smooth functions (also known as Hölder smooth functions with exponent \( \beta \)) as

\[
\Phi(\beta, c) = \left\{ h : \sup_{\|a\| \leq \beta} |D^a h(x)| \leq c, \sup_{\|a\| = \beta} \sup_{x \in X} \frac{|D^a h(x) - D^a h(y)|}{\|x - y\|_2^{\beta - \|a\|}} \leq c \right\},
\]

for some constant \( c > 0 \), where \( \|\beta\| \) denotes the largest integer that is smaller than \( \beta \) and \( D^a \) denotes the differential operator \( D^a \) denote the differential operator:

\[
D^a h(x) = \frac{\partial^{\|a\|}}{\partial x_1^{a_1} \cdots \partial x_d^{a_d}} h(x).
\]

We introduce the following conditions.

(A6.) Suppose \( Q(\cdot, a) \in \Phi(\beta, c) \), and \( p(a|\cdot) \in \Phi(\beta, c) \) for any \( a \).

(A7.) Functions \( \{\hat{q}_\mathcal{I} \}_{\mathcal{I} \in \mathcal{P}} \) are uniformly bounded.

Assumption (A7) ensures that the optimizer would not diverge in the \( \ell_\infty \) sense. Similar assumptions are commonly imposed in the literature to derive the convergence rates of DNN estimators (see e.g., Farrell et al. 2021). Combining (A7) with (A6) allows us to derive the uniform rate of convergence for the class of DNN estimators \( \{\hat{q}_\mathcal{I} \}_{\mathcal{I} \in \mathcal{P}} \). The following theorem summarizes the theoretical properties of the proposed method via deep neural networks.

**Theorem 3** Assume (A1)-(A3), (A6), (A7) and Model I hold. Assume \( X \) and \( Y \) are bounded variables, and \( A \) has a bounded probability density function on \([0, 1]\). Assume \( m \asymp n, \{\gamma_n \}_{n \in \mathbb{N}} \) satisfies \( \gamma_n \to 0 \) and \( \gamma_n \gg n^{-2p/(2\beta + p)} \log^8 n \). Then, there exist some constant \( \bar{c} > 0 \) and DNN classes \( \{\mathcal{Q}_\mathcal{I} : \mathcal{I} \} \) with \( L_\mathcal{I} \asymp \log(n|\mathcal{I}|) \) and \( W_\mathcal{I} \asymp (n|\mathcal{I}|)^{p/(2\beta + p)} \log(n|\mathcal{I}|) \) such that the resulting D-JIL estimator computed by (2.3) satisfies

(i) \( |\hat{\mathcal{S}}| = |\mathcal{P}_0| \).
(ii) $\max_{\tau \in J} \min_{\hat{\tau} \in \hat{J}} |\hat{\tau} - \tau| \leq \hat{\epsilon} n^{-2\beta/(2\beta + p)} \log^8 n$;

(iii) $E|Q(X, A) - \sum_{\theta \in \Theta} \mathbb{I}(A \in \theta) \hat{\theta}(X)|^2 dA \leq \hat{\epsilon} n^{-2\beta/(2\beta + p)} \log^8 n$,

with probability at least $1 - O(n^{-2})$.

Theorem 3 establishes the properties of our method under settings where the $Q(x, a)$ is piecewise function in the treatment. Results in (i) imply that D-JIL correctly identifies the number of change points. Results in (ii) imply that any change point in $P_0$ can be consistently identified at a convergence rate of $O_p(n^{-2\beta/(2\beta + p)})$ up to some logarithmic factors. Notice that we use the piecewise function $\sum_{\theta \in \Theta} \mathbb{I}(A \in \theta) \hat{\theta}(x)$ to approximate the outcome regression function. In (iii), we show our estimator for function $Q(X, A)$ converges at a rate of $O_p(n^{-2\beta/(2\beta + p)})$ up to some logarithmic factors. The theoretical choices of $L_\theta$ and $W_\theta$ in Theorem 3 are consistent with the literature of DNN estimators (Imaizumi and Fukumizu 2019; Farrell et al. 2021). These DNN architectures ensure the convergence rate of our estimator for function $Q(X, A)$, which achieves the minimax-optimal nonparametric rate of convergence under (A6) (see e.g., Stone 1982).

We next establish the convergence rate of $V^{opt} - V(\hat{d})$ under Model I in the following theorem.

**Theorem 4** Assume the conditions in Theorem 3 are satisfied. Further assume (A5) holds. Then, we have

$$V(\hat{d}) \geq V^{opt} - O(1)(n^{-2\beta/(2\beta + p)} \log^8 n + n^{-2\beta(1+\gamma)/(2\beta + p) + \gamma}) \log^{8+\gamma} n),$$

(2.12)

with probability at least $1 - O(n^{-2})$.

Theorem 4 suggests that $V(\hat{d})$ converges to the optimal value at a rate of $O_p\{n^{-2\beta/(2\beta + p)} \log^8 n + n^{-2\beta(1+\gamma)/(2\beta + p) + \gamma} \log^{8+\gamma} n\}$ up to some logarithmic factors. This rate is slower than the rate ($O_p(n^{-1})$ up to some logarithmic factor) we obtained in Theorem 2 where we posit a parametric (linear) model. Suppose the condition $4\beta(1 + \gamma) > (2\beta + p)(2 + \gamma)$ holds, it follows that $V(\hat{d}) = V^{opt} + o_p(n^{-1/2})$. This observation forms the basis of our inference procedure in Section 2.4.1. Here, the extra margin parameter $\gamma$ in our results is introduced by (A5) to bound the bias due to the estimated decision rule $\hat{d}$. If the margin parameter $\gamma$ goes to infinity, we only require the smooth parameter $\beta > p/2$ to obtain $V(\hat{d}) = V^{opt} + o_p(n^{-1/2})$. This condition ($\beta > p/2$) is commonly assumed in the literature on evaluating average treatment effects (see e.g., Chernozhukov et al. 2017; Farrell et al. 2021).
Evaluation of the Value Function

Suppose Model I holds. When L-JIL is used, it follows from Theorem 2 that \( V(\hat{d}) = V^{opt} + o_p(n^{-1/2}) \). When D-JIL is used, if the smoothness parameter \( \beta \) (see (A6)) and the margin parameter \( \gamma \) (see (A5)) satisfy \( 4\beta(1+\gamma) > (2\beta+p)(2+\gamma) \), it follows from Theorem 4 that \( V(\hat{d}) = V^{opt} + o_p(n^{-1/2}) \). In the following, we derive the asymptotic normality of \( \sqrt{n}(\hat{V} - V^{opt}) \). By Slutsky’s theorem, this implies that \( \sqrt{n}(\hat{V} - V(\hat{d})) \) is asymptotically normal as well.

\[
\begin{align*}
(A8.) \quad & \mathbb{E}\left[\left(\hat{\beta}(\mathcal{S}|X) - \beta(\mathcal{S}|X)\right)^2\right]^{1/2} = o(n^{-1/4}) \quad \text{and that} \quad \hat{\beta}(\mathcal{S};\bullet) \quad \text{belongs to the class of VC-type functions with VC-index upper bounded by} \quad O(n^{1/2}),
\end{align*}
\]

for any \( \mathcal{S} \in \mathcal{J}(m) \).

The first part in Assumption (A8) requires the generalized propensity score function to converge at certain rates. Similar assumptions are commonly imposed in the causal inference literature to derive the asymptotic distribution of the estimated average treatment effect (see e.g., Chernozhukov et al. 2017). The second part of (A8) essentially controls the model complexity of the estimator \( \hat{\beta} \). The more complicated \( \hat{\beta} \) is, the larger the VC index. Under (A6), we can show (A8) holds when DNN is used to model the generalized propensity score.

**Theorem 5** Assume (A8) holds and suppose functions \( \{\hat{\beta}(\mathcal{S})\}_{\mathcal{S} \in \mathcal{F}} \) are uniformly bounded away from zero. Further assume that for any \( \mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}_0 \) with \( \mathcal{S}_1 \neq \mathcal{S}_2 \), we have \( \Pr(q_{\mathcal{S}_1,0}(X) = q_{\mathcal{S}_2,0}(X)) = 0 \).

(i) Suppose conditions in Theorem 2 are satisfied. Then, under L-JIL, we have
\[
\sqrt{n}(\hat{V} - V^{opt}) \overset{d}{\rightarrow} N(0, \sigma^2_L),
\]
for some \( \sigma^2_L > 0 \).

(ii) Suppose conditions in Theorem 4 are satisfied with \( 4\beta(1+\gamma) > (2\beta+p)(2+\gamma) \). Then, under D-JIL, we have
\[
\sqrt{n}(\hat{V} - V^{opt}) \overset{d}{\rightarrow} N(0, \sigma^2_D),
\]
for some \( \sigma^2_D > 0 \).

We now introduce the estimator for the asymptotic variance \( \sigma^2_L \) or \( \sigma^2_D \), and derive a Wald-type \( 1-\alpha \) CI for \( V^{opt} \). Since \( V(\hat{d}) = V^{opt} + o_p(n^{-1/2}) \), the proposed CI also covers \( V(\hat{d}) \) with probability tending to \( 1-\alpha \). We estimate \( \sigma^2_L \) or \( \sigma^2_D \) by
\[
\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[ \frac{\Pi_{i \in \hat{d}(X_i)}}{\hat{\beta}(\hat{d}(X_i)|X_i)} \left( Y_i - \max_{\mathcal{S} \in \mathcal{F}} \hat{q}_\mathcal{S}(X_i) \right) + \max_{\mathcal{S} \in \mathcal{F}} \hat{q}_\mathcal{S}(X_i) - \hat{V} \right]^2,
\]

25
where \{\hat{q}_\theta(\cdot)\} corresponds to the value estimations under L-JIL or D-JIL.

The corresponding $1 - \alpha$ CI is given by $\hat{V} \pm z_{\alpha/2} \hat{\sigma}$, where $z_{\alpha/2}$ denotes the upper $\alpha/2$-th quantile of a standard normal distribution. Similar to Theorem 5, we can show that $\hat{\sigma}$ is consistent. This shows the validity of our inference procedure.

### 2.4.2 Properties under Model II

#### Properties of L-JIL under Varying Coefficient Model

We first consider the case when the outcome regression function can be represented by a varying coefficient model and investigate the theoretical properties of the proposed L-JIL. Specifically, suppose the true outcome regression function takes the following form

$$Q(x, a) = \bar{x}^T \theta_0(a), \quad \forall x \in \mathbb{X}, a \in [0, 1], \quad (2.13)$$

where $\bar{x} = (1, x^T)^T$ and $\theta_0(\cdot)$ is some continuous $(p + 1)$-dimensional function. That is, we assume the conditional mean of the outcome is a linear function of individuals’ covariates for any treatment $a \in [0, 1]$. Yet, the model is flexible in that $\theta_0(\cdot)$ is allowed to be either a step function, or an arbitrary continuous function of $a$ with certain smoothness constraints. Models of this type belong to the class of varying coefficient models popularly applied in many scientific areas (see e.g., Fan and Zhang 2008, for an overview).

Here, we consider the following class of Hölder continuous functions for $\theta_0(\cdot)$. Suppose there exist some constants $L > 0$, $0 < \alpha_0 \leq 1$ such that $\theta_0(\cdot)$ satisfies

$$\sup_{a_1, a_2 \in [0, 1]} \|\theta_0(a_1) - \theta_0(a_2)\|_2 \leq L|a_1 - a_2|^\alpha_0. \quad (2.14)$$

We first sketch a few lines to see why our method works under (2.14). For a given integer $k > 0$, we define $\theta_k^*(\cdot)$ as

$$\theta_k^*(a) = \sum_{j=0}^{k-1} \theta_0\left(\frac{j + 1/2}{k + 1}\right) \mathbb{I}(j \leq (k + 1)a < j + 1) + \theta_0\left(\frac{k + 1/2}{k + 1}\right) \mathbb{I}((k + 1)a \geq k).$$

Apparently, $\theta_k^*(\cdot)$ has at most $k$ change points. In addition, with some calculations, we can show that $\sup_{a \in [0, 1]} \|\theta_k^*(a) - \theta_0(a)\|_2 \leq 2^{-\alpha_0}(k + 1)^{-\alpha_0} L$. Letting $k \to \infty$, it is immediate to see that $\theta_0(\cdot)$ can be uniformly approximately by a step function as the number of change points increases.
In Theorems 1 and 2, we have shown the proposed I2DR is consistent under the piecewise linear function assumption. Based on the above discussion, we expect that jump interval-learning also works when the model (2.13) holds. We formally establish the corresponding theoretical results in the following theorem.

**Theorem 6** Assume (A1)-(A4) and (2.14) hold. Assume A has a bounded probability on $[0, 1]$. Assume $m \asymp n$, $\lambda_n = O(n^{-1}\log n)$, $\gamma_n$ satisfies $\gamma_n \to 0$ and $\gamma_n \gg n^{-1}\log n$. Under the model (2.13), there exists some constant $\bar{c} > 0$ such that the following holds with probability at least $1 - O(n^{-2})$:

$$
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2 \, da \leq \bar{c} \gamma_n^{2\alpha_0/(1+2\alpha_0)}.
$$

In addition, assume $\gamma_n \asymp (n^{-1}\log n)^{(1+2\alpha_0)/(1+4\alpha_0)}$. Then there exists some constant $\bar{c}^* > 0$ such that the following occurs with probability at least $1 - O(n^{-2})$ that

$$
V^{opt} - V(\hat{d}) \leq \bar{c}^*(n^{-1}\log n)^{\alpha_0/(1+4\alpha_0)}.
$$

(2.15)

It is worth mentioning that with proper choice of $\gamma_n$, the integrated $\ell_2$ loss of $\hat{\theta}(\cdot)$ converges at a rate of $O_p(n^{-2\alpha_0/(1+2\alpha_0)})$ up to some logarithmic factor. The rate is slower compared to the results in Theorem 1, since $\theta_0(\cdot)$ is only "approximately" piecewise constant. When $\theta_0(\cdot)$ is Lipschitz continuous, it follows from (2.15) that the value under our proposed I2DR will converge to the optimal value function at a rate of $O_p(n^{-1/5}\log^{1/5} n)$.

**Properties of D-JIL under the Continuous Outcome Regression Function**

We next consider the general case when the outcome regression function is specified by model II and study the theoretical properties of the proposed D-JIL. The following theorem proves the consistency of the proposed estimator.

**Theorem 7** Suppose $Q$ is a continuous function of $a$ and $x$. Assume (A1)-(A3) and (A6)-(A7) hold. Assume $X$ and $Y$ are bounded variables, and A has a bounded probability density function on $[0, 1]$. Assume $m \asymp n$ and $\{\gamma_n\}_{n \in \mathbb{N}}$ satisfies $\gamma_n \to 0$ and $\gamma_n \gg n^{-2\beta/(2\beta+p)}\log^8 n$. Then, there exist some DNN classes $\{\mathcal{Q}_I: \mathcal{J}\}$ with $L_\mathcal{Q}_I \asymp \log(n|\mathcal{J}|)$ and $W_\mathcal{Q}_I \asymp (n|\mathcal{J}|)^{p/(2\beta+p)}\log(n|\mathcal{J}|)$ such that the resulting D-JIL estimator computed by (2.3) satisfies

(i) $\max_{\mathcal{Q} \in \mathcal{Q}_I} \sup_{a \in \mathcal{J}} E[\bar{q}_{\mathcal{Q}}(X) - Q(X, a)]^2 = o_p(1)$;

(ii) $V^{opt} - V(\hat{d}) = o_p(1)$. 

27
Theorem 7 establishes the properties of our method under settings where $Q$ is continuous in $a$. Results in (i) imply that $\tilde{q}_\mathcal{F}(x)$ can be used to uniformly approximate $Q(x, a)$ for any $a \in \mathcal{F}$. The consistency of the value in (ii) thus follows.

2.5 Simulations

2.5.1 Confidence Interval for the Value

In this section, we focus on scenarios where the outcome regression function takes the form of Model I and examine the coverage probability of the proposed CI in Section 2.4.1. Simulated data are generated from the following model:

$$Y|X, A \sim N(Q(X, A), 1), \ A|X \sim \text{Unif}[0, 1] \text{ and } X^{(1)}, X^{(2)}, \ldots, X^{(p)} \overset{iid}{\sim} N(0, 1)$$

where $\text{Unif}[a, b]$ denotes the uniform distribution on the interval $[a, b]$. We consider the following two scenarios with different choices of $Q(X, A)$.

**Scenario 1:**

$$Q(x, a) = \begin{cases} 
1 + x^{(1)}, & a < 0.35, \\
x^{(1)} - x^{(2)}, & 0.35 \leq a < 0.65, \\
1 - x^{(2)}, & a \geq 0.65.
\end{cases}$$

Under Scenario 1, the outcome regression function is piecewise constant as a function of $a$, and is linear as a function of $x$. Here, we have $J(\mathcal{P}_0) = \{0.35, 0.65\}$ and $|\mathcal{P}_0| = 3$. With some calculations, one can show that the optimal value $V^{opt}$ equals 1.34.

**Scenario 2:**

$$Q(x, a) = \begin{cases} 
1 + (x^{(1)})^3, & a < 0.35, \\
x^{(1)} - \log(1.5 + x^{(2)}), & 0.35 \leq a < 0.65, \\
1 - \sin(0.5\pi x^{(2)}), & a \geq 0.65.
\end{cases}$$

Under Scenario 2, the outcome regression function is piecewise constant as a function of $a$, but is nonlinear as a function of $x$. The change points are $J(\mathcal{P}_0) = \{0.35, 0.65\}$ with $|\mathcal{P}_0| = 3$. The optimal value equals 1.35, based on Monte Carlo approximations.

For each scenario, we set $p = 4$ and consider three different choices of the sample size, corresponding to $n = 200, 400, 800$. We apply the proposed D-JIL to both scenarios. L-JIL is
Table 2.1: The estimated optimal value $\hat{V}$ with its standard error, the empirical coverage probability of its associated confidence interval, and the averaged number of estimated partitions computed by the proposed L-JIL and D-JIL.

<table>
<thead>
<tr>
<th>Method</th>
<th>Optimal value $V_{opt}$</th>
<th>Scenario 1, $p = 4$</th>
<th>Scenario 2, $p = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
</tr>
<tr>
<td>L-JIL</td>
<td>Estimated optimal value $\hat{V}$</td>
<td>1.436</td>
<td>1.383</td>
</tr>
<tr>
<td></td>
<td>Mean of standard error $\hat{\sigma}$</td>
<td>0.129</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>Coverage probabilities(%)</td>
<td>89.80</td>
<td>93.20</td>
</tr>
<tr>
<td></td>
<td>Number of partitions $</td>
<td>\mathcal{P}</td>
<td>$</td>
</tr>
<tr>
<td>D-JIL</td>
<td>Estimated optimal value $\hat{V}$</td>
<td>1.297</td>
<td>1.338</td>
</tr>
<tr>
<td></td>
<td>Mean of standard error $\hat{\sigma}$</td>
<td>0.160</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>Coverage probabilities(%)</td>
<td>90.60</td>
<td>93.60</td>
</tr>
<tr>
<td></td>
<td>Number of partitions $</td>
<td>\mathcal{P}</td>
<td>$</td>
</tr>
</tbody>
</table>

applied to Scenario 1 only, as it requires the outcome regression function to be linear in the baseline covariates. The detailed implementation is discussed in Section 2.3.3. We set $m = n/5$, $\lambda_n = 0$, $\gamma_n = 4n^{-1}\log(n)$, and construct the CI for $V_{opt}$ based on the procedure described in Section 2.4.1. Reported in Table 1 are the estimated value function $\hat{V}$ with its standard error $\hat{\sigma}$, the empirical coverage probabilities of the proposed confidence interval for $V_{opt}$, and the number of estimated partitions $|\mathcal{P}|$, aggregated over 500 simulations.

Based on the results, it is clear that the estimated value function approaches the optimal value as the sample size increases for both two methods. For instance, when $n = 800$, L-JIL obtained an estimated value of 1.340 under Scenario 1 on average. D-JIL yields an average value of 1.349 under Scenario 2. These values are very close to the truths 1.34 and 1.35, respectively. The performance of our proposed L-JIL and D-JIL are comparable under the Scenario 1. In addition, as the sample size increases, the coverage probability of the Wald-type CI approaches to the nominal level. This verifies our theoretical findings in Theorem 5. Moreover, the averaged estimated number of partitions $|\mathcal{P}|$ is approximately 3 for all settings. This supports our theoretical findings in Theorems 1 and 3.

2.5.2 Value Function under the Proposed I2DR

In this section, we consider more general settings and compare the proposed procedure with the existing state-of-the-art methods that outputs single-valued decision rule. Similar
to Section 2.5.1, we generate the data from the following model:

\[ Y \mid X, A \sim N(Q(X, A), 1), \ A \mid X \sim \text{Unif}[0, 1] \text{ and } X^{(1)}, X^{(2)}, \ldots, X^{(p)} \overset{iid}{\sim} N(0, 1). \]

In addition to Scenarios 1 and 2, we consider several other choices of the outcome regression function, allowing the working model assumption in Model I or Model II to be violated in some scenarios.

**Scenario 3:**

\[
Q(x, a) = \begin{cases} 
\sqrt{x^{(1)}/2 + 0.5}, & a < 0.25, \\
\sin(2\pi x^{(2)}), & 0.25 \leq a < 0.5, \\
0.5 - (x^{(1)} + x^{(2)} - 0.75)^2, & 0.5 \leq a < 0.75, \\
0.5, & a \geq 0.75.
\end{cases}
\]

The outcome regression function in Scenario 3 is piecewise function of \( a \) but nonlinear function of \( x \) with complex treatment-covariates interactions. We have \( J(\mathcal{P}_0) = \{0.25, 0.5, 0.75, 1\} \) with \( |\mathcal{P}_0| = 4 \), and the optimal value equals 0.76.

**Scenario 4:**

\[
Q(x, a) = \bar{x}^\top \{2|a - 0.5|\theta^*\},
\]

where \( \theta^* = (1, 2, -2, 0_{p-2})^\top \). By setting \( \theta_0(a) = 2|a-0.5|\theta^* \), it is immediate to see that \( Q(x, a) = \bar{x}^\top \theta_0(a) \) and satisfies the condition in (2.13). Note that \( \theta_0(\cdot) \) here is a continuous function. One can show that \( V^{opt} = 1.28 \) under this scenario.

**Scenario 5:**

\[ Q(x, a) = 8 + 4x^{(1)} - 2x^{(2)} - 2x^{(3)} - 10(1 + 0.5x^{(1)} + 0.5x^{(2)} - 2a)^2. \]

This scenario is considered in Chen et al. (2016). Note that the outcome regression function is continuous in both the baseline covariate and the treatment. With some calculations, one can show that \( V^{opt} = 8 \).

We apply the proposed L-JIL and D-JIL to estimate the optimal I2DR for Scenario 1-5, with \( p = 20 \) and \( n \in \{50, 100, 200, 400, 800\} \). The tuning parameters in JILs are specified according to Section 2.3.4. Here, we set \( m = n/c \) with \( c = 10 \) to save computational cost. In Section 2.5.3, we report results with \( c \in \{6, 8\} \) and find the values under the estimated I2DRs are very similar to those with \( c = 10 \).

To evaluate the proposed I2DRs, we compare its value function \( V(\hat{d}) \) with the values
Table 2.2: The value function under the proposed I2DR and IDR s estimated based on outcome weighted learning for Scenario 1-5.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L-JIL</td>
<td>0.783(0.016)</td>
<td>0.832(0.016)</td>
<td>1.080(0.014)</td>
<td>1.259(0.002)</td>
<td>1.297(0.001)</td>
</tr>
<tr>
<td></td>
<td>D-JIL</td>
<td>0.914(0.012)</td>
<td>0.967(0.008)</td>
<td>1.050(0.005)</td>
<td>1.071(0.005)</td>
<td>1.138(0.001)</td>
</tr>
<tr>
<td></td>
<td>V = 1.34</td>
<td>p = 20</td>
<td>L-O-L</td>
<td>0.558(0.004)</td>
<td>0.574(0.004)</td>
<td>0.600(0.005)</td>
</tr>
<tr>
<td></td>
<td>K-O-L</td>
<td>0.335(0.008)</td>
<td>0.415(0.006)</td>
<td>0.441(0.006)</td>
<td>0.457(0.005)</td>
<td>0.489(0.004)</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>L-JIL</td>
<td>0.741(0.021)</td>
<td>0.854(0.020)</td>
<td>1.180(0.007)</td>
<td>1.266(0.001)</td>
<td>1.299(0.001)</td>
</tr>
<tr>
<td></td>
<td>D-JIL</td>
<td>0.900(0.012)</td>
<td>0.978(0.008)</td>
<td>1.074(0.004)</td>
<td>1.102(0.003)</td>
<td>1.141(0.001)</td>
</tr>
<tr>
<td></td>
<td>V = 1.35</td>
<td>p = 20</td>
<td>L-O-L</td>
<td>0.450(0.009)</td>
<td>0.448(0.006)</td>
<td>0.447(0.005)</td>
</tr>
<tr>
<td></td>
<td>K-O-L</td>
<td>0.115(0.019)</td>
<td>0.213(0.010)</td>
<td>0.229(0.007)</td>
<td>0.241(0.004)</td>
<td>0.276(0.002)</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>L-JIL</td>
<td>0.227(0.020)</td>
<td>0.268(0.013)</td>
<td>0.372(0.008)</td>
<td>0.432(0.003)</td>
<td>0.511(0.002)</td>
</tr>
<tr>
<td></td>
<td>D-JIL</td>
<td>0.453(0.019)</td>
<td>0.469(0.009)</td>
<td>0.511(0.005)</td>
<td>0.526(0.004)</td>
<td>0.545(0.002)</td>
</tr>
<tr>
<td></td>
<td>V = 0.76</td>
<td>p = 20</td>
<td>L-O-L</td>
<td>0.002(0.010)</td>
<td>-0.009(0.008)</td>
<td>-0.060(0.006)</td>
</tr>
<tr>
<td></td>
<td>K-O-L</td>
<td>-0.268(0.026)</td>
<td>-0.233(0.015)</td>
<td>-0.260(0.009)</td>
<td>-0.251(0.006)</td>
<td>-0.233(0.003)</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>L-JIL</td>
<td>0.553(0.013)</td>
<td>0.564(0.011)</td>
<td>0.630(0.011)</td>
<td>0.806(0.006)</td>
<td>0.882(0.002)</td>
</tr>
<tr>
<td></td>
<td>D-JIL</td>
<td>0.612(0.014)</td>
<td>0.651(0.008)</td>
<td>0.684(0.004)</td>
<td>0.653(0.006)</td>
<td>0.801(0.001)</td>
</tr>
<tr>
<td></td>
<td>V = 1.28</td>
<td>p = 20</td>
<td>L-O-L</td>
<td>0.525(0.016)</td>
<td>0.458(0.010)</td>
<td>0.375(0.004)</td>
</tr>
<tr>
<td></td>
<td>K-O-L</td>
<td>0.236(0.007)</td>
<td>0.260(0.004)</td>
<td>0.252(0.003)</td>
<td>0.244(0.001)</td>
<td>0.246(0.001)</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>L-JIL</td>
<td>5.82(0.05)</td>
<td>6.41(0.02)</td>
<td>6.80(0.01)</td>
<td>7.02(0.01)</td>
<td>7.16(0.01)</td>
</tr>
<tr>
<td></td>
<td>D-JIL</td>
<td>5.57(0.06)</td>
<td>5.79(0.03)</td>
<td>5.97(0.02)</td>
<td>6.10(0.01)</td>
<td>6.26(0.01)</td>
</tr>
<tr>
<td></td>
<td>V = 8.00</td>
<td>p = 20</td>
<td>L-O-L</td>
<td>5.92(0.07)</td>
<td>6.75(0.03)</td>
<td>7.32(0.02)</td>
</tr>
<tr>
<td></td>
<td>K-O-L</td>
<td>6.70(0.02)</td>
<td>7.05(0.02)</td>
<td>7.38(0.01)</td>
<td>7.58(0.01)</td>
<td>7.56(0.01)</td>
</tr>
</tbody>
</table>

under estimated optimal IDRs obtained by linear outcome weighted learning (L-O-L) and the nonlinear outcome weighted learning based on the Gaussian kernel function (K-O-L). Both methods were proposed by Chen et al. (2016). To implement L-O-L and K-O-L, we fix the parameter \( \phi_n = 0.1 \), and select other tuning parameters by five-fold cross-validation, as in Chen et al. (2016). All the value functions are evaluated via Monte Carlo simulations. The average value function as well as its standard deviation over 200 replicates are summarized in Table 2.2.

It can be seen from Table 2 that both L-JIL and D-JIL are very efficient when Model I (Scenarios 1-3) holds, and perform reasonably well when Model II (Scenario 4 and 5) holds or the sample size is small. For instance, the proposed L-JIL achieves a value of 1.297 in Scenario 1 and 1.299 in Scenario 2, when \( n = 800 \). These values are very close to the optimal values, given by 1.34 and 1.35. In Scenario 3, the proposed D-JIL performs consistently better than L-JIL, due to the capacity of deep neural networks in approximating complicated non-linear relationships. In addition, it can be seen that D-JIL achieves the best performance with small sample size in most cases. Moreover, the value of the proposed
Table 2.3: The value function of the proposed I2DR under L-JIL for Scenario 1-5 with different choice of $m = n/c$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>$c = 6$</td>
<td>0.813(0.019)</td>
<td>0.858(0.017)</td>
<td>1.027(0.014)</td>
<td>1.249(0.003)</td>
<td>1.289(0.001)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.836(0.022)</td>
<td>0.870(0.018)</td>
<td>1.024(0.014)</td>
<td>1.238(0.002)</td>
<td>1.295(0.001)</td>
</tr>
<tr>
<td></td>
<td>$c = 10$</td>
<td>0.783(0.016)</td>
<td>0.832(0.016)</td>
<td>1.080(0.014)</td>
<td>1.259(0.002)</td>
<td>1.297(0.001)</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>$c = 6$</td>
<td>0.804(0.025)</td>
<td>0.891(0.021)</td>
<td>1.132(0.008)</td>
<td>1.257(0.002)</td>
<td>1.290(0.001)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.857(0.029)</td>
<td>0.935(0.021)</td>
<td>1.123(0.009)</td>
<td>1.241(0.002)</td>
<td>1.299(0.001)</td>
</tr>
<tr>
<td></td>
<td>$c = 10$</td>
<td>0.741(0.021)</td>
<td>0.854(0.020)</td>
<td>1.180(0.007)</td>
<td>1.266(0.001)</td>
<td>1.299(0.001)</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>$c = 6$</td>
<td>0.280(0.023)</td>
<td>0.310(0.014)</td>
<td>0.339(0.008)</td>
<td>0.422(0.003)</td>
<td>0.504(0.002)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.229(0.019)</td>
<td>0.325(0.014)</td>
<td>0.326(0.008)</td>
<td>0.417(0.003)</td>
<td>0.512(0.002)</td>
</tr>
<tr>
<td></td>
<td>$c = 10$</td>
<td>0.227(0.020)</td>
<td>0.268(0.013)</td>
<td>0.372(0.008)</td>
<td>0.432(0.003)</td>
<td>0.511(0.002)</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>$c = 6$</td>
<td>0.565(0.015)</td>
<td>0.561(0.012)</td>
<td>0.639(0.011)</td>
<td>0.818(0.006)</td>
<td>0.884(0.002)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.563(0.015)</td>
<td>0.564(0.012)</td>
<td>0.627(0.011)</td>
<td>0.810(0.006)</td>
<td>0.882(0.002)</td>
</tr>
<tr>
<td></td>
<td>$c = 10$</td>
<td>0.553(0.013)</td>
<td>0.564(0.011)</td>
<td>0.630(0.011)</td>
<td>0.806(0.006)</td>
<td>0.882(0.002)</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>$c = 6$</td>
<td>5.81(0.05)</td>
<td>6.38(0.02)</td>
<td>6.78(0.01)</td>
<td>6.99(0.01)</td>
<td>7.09(0.01)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>5.82(0.05)</td>
<td>6.40(0.02)</td>
<td>6.78(0.01)</td>
<td>7.02(0.01)</td>
<td>7.12(0.01)</td>
</tr>
<tr>
<td></td>
<td>$c = 10$</td>
<td>5.82(0.05)</td>
<td>6.41(0.02)</td>
<td>6.80(0.01)</td>
<td>7.02(0.01)</td>
<td>7.16(0.01)</td>
</tr>
</tbody>
</table>

I2DR increases with the sample size in most cases. This supports our theoretical findings in Section 2.4.

In comparison, the value function under the estimated IDR using L-O-L or K-O-L is no more than half of the optimal value, for each setting in Scenario 1 to 3. In Scenario 4, both L-JIL and D-JIL achieve much larger value functions than outcome weighted learning. In Scenario 5, L-O-L and K-O-L have better performance, as the true optimal decision rule is linear and the outcome regression function is very sensitive to the change of the treatment level $a$ (by noticing that the coefficient of the quadratic term in Scenario 5 is 10). When $n = 800$, the two competing methods achieve similar value functions.

### 2.5.3 Choice of $m$

Recall that we set $n = m/c$ for some constant $c > 0$. In Sections 2.5.2, we report our simulation results with $c = 10$. In this section, we set $c \in \{6, 8\}$ and report the corresponding results in Tables 2.3 and 2.4 for L-JIL and D-JIL, respectively. We also include results with $c = 10$ for completeness. It can be seen that the value functions are very similar across different choices of $c$.
Table 2.4: The value function of the proposed I2DR under D-JIL for Scenario 1-5 with different choice of $m = n/c$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 1</td>
<td>$V = 1.34$</td>
<td>$c = 6$</td>
<td>0.941(0.012)</td>
<td>0.972(0.008)</td>
<td>1.028(0.004)</td>
<td>1.065(0.004)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.973(0.016)</td>
<td>0.990(0.008)</td>
<td>1.030(0.004)</td>
<td>1.053(0.005)</td>
<td>1.136(0.001)</td>
</tr>
<tr>
<td></td>
<td>$p = 20$</td>
<td>$c = 10$</td>
<td>0.914(0.012)</td>
<td>0.967(0.008)</td>
<td>1.050(0.005)</td>
<td>1.071(0.005)</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>$V = 1.35$</td>
<td>$c = 6$</td>
<td>0.943(0.013)</td>
<td>0.980(0.008)</td>
<td>1.037(0.004)</td>
<td>1.087(0.003)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>1.002(0.015)</td>
<td>1.012(0.008)</td>
<td>1.039(0.004)</td>
<td>1.076(0.003)</td>
<td>1.137(0.001)</td>
</tr>
<tr>
<td></td>
<td>$p = 20$</td>
<td>$c = 10$</td>
<td>0.900(0.012)</td>
<td>0.978(0.008)</td>
<td>1.074(0.004)</td>
<td>1.102(0.003)</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>$V = 0.76$</td>
<td>$c = 6$</td>
<td>0.475(0.018)</td>
<td>0.480(0.009)</td>
<td>0.481(0.006)</td>
<td>0.493(0.004)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.416(0.019)</td>
<td>0.497(0.009)</td>
<td>0.493(0.006)</td>
<td>0.506(0.003)</td>
<td>0.532(0.002)</td>
</tr>
<tr>
<td></td>
<td>$p = 20$</td>
<td>$c = 10$</td>
<td>0.453(0.019)</td>
<td>0.469(0.009)</td>
<td>0.511(0.005)</td>
<td>0.526(0.004)</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>$V = 1.28$</td>
<td>$c = 6$</td>
<td>0.624(0.014)</td>
<td>0.655(0.008)</td>
<td>0.686(0.004)</td>
<td>0.687(0.005)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>0.622(0.014)</td>
<td>0.651(0.008)</td>
<td>0.684(0.004)</td>
<td>0.676(0.005)</td>
<td>0.801(0.001)</td>
</tr>
<tr>
<td></td>
<td>$p = 20$</td>
<td>$c = 10$</td>
<td>0.612(0.014)</td>
<td>0.651(0.008)</td>
<td>0.684(0.004)</td>
<td>0.653(0.006)</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>$V = 8.00$</td>
<td>$c = 6$</td>
<td>5.49(0.06)</td>
<td>5.69(0.03)</td>
<td>5.82(0.02)</td>
<td>5.97(0.01)</td>
</tr>
<tr>
<td></td>
<td>$c = 8$</td>
<td>5.58(0.05)</td>
<td>5.77(0.03)</td>
<td>5.91(0.02)</td>
<td>6.04(0.01)</td>
<td>6.20(0.01)</td>
</tr>
<tr>
<td></td>
<td>$p = 20$</td>
<td>$c = 10$</td>
<td>5.57(0.06)</td>
<td>5.79(0.03)</td>
<td>5.97(0.02)</td>
<td>6.10(0.01)</td>
</tr>
</tbody>
</table>

2.6 Real Data Analysis

In this section, we illustrate the empirical performance of our proposed method on a real data from the International Warfarin Pharmacogenetics Consortium. Warfarin is commonly used for preventing thrombosis and thromboembolism. High doses of Warfarin are more beneficial than its lower doses, but may lead to a high risk of bleeding as well. Proper dosing of Warfarin is thus of significant importance.

We use the dataset provided by the International Warfarin Pharmacogenetics Consortium (2009) for analysis. We choose 6 baseline covariates, including age, height, weight, gender, the VKORC1.AG genotype and VKORC1.AA genotype. This yields a total of 3848 with complete records of baseline information. The outcome is defined as the absolute distance between the international normalized ratio (INR, a measurement of the time it takes for the blood to clot) after the treatment and the ideal value 2.5, i.e, $Y = -|\text{INR} - 2.5|$. We use the min-max normalization to convert the range of the dose level $A$ into $[0, 1]$.

To implement L-JIL and D-JIL, we set $m = n/5$, and select $\gamma_n$ and $\lambda_n$ via cross validation, as in Section 2.5.2. To further evaluate the empirical performance of the proposed I2DRs, we compare their values with the value under the IDR estimated by K-O-L. Specifically, we randomly select 70% of the data to compute the proposed I2DR and the IDR obtained by
Table 2.5: The regression coefficients associated with three subintervals for the proposed I2DR under L-JIL.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Age</th>
<th>Weight</th>
<th>Height</th>
<th>Gender</th>
<th>VKORC1.AG</th>
<th>VKORC1.AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>-1.673</td>
<td>0.025</td>
<td>0.006</td>
<td>0.006</td>
<td>-0.158</td>
<td>-0.364</td>
<td>-0.349</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-1.741</td>
<td>0.029</td>
<td>0.004</td>
<td>0.006</td>
<td>-0.201</td>
<td>0.057</td>
<td>-0.051</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>-0.488</td>
<td>0.012</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.033</td>
<td>-0.002</td>
<td>-0.120</td>
</tr>
</tbody>
</table>

K-O-L, and evaluate their value functions using the remaining dataset. We then iterate this procedure 50 times to calculate the average value function. For each iteration, the value function is estimated based on the nonparametric estimator proposed by Zhu et al. (2020a).

Specifically, let $G_{test}$ denote observations in the testing dataset. For the IDR $\tilde{d}$ computed by K-O-L, we consider the following nonparametric estimator for its value function,

$$
\tilde{V}(\tilde{d}) = \int \sum_{i \in G_{test}} \frac{Y_i K(h_x^{-1}(x - X_i))K(h_a^{-1}(\tilde{d}(x) - A_i))}{|G_{test}|h_x^p} \left\{ \sum_{i \in G_{test}} \frac{K(h_x^{-1}(x - X_i))}{|G_{test}|h_x^p} \right\} dx,
$$

where $K(\cdot)$ denotes the Gaussian kernel function, and $h_x$ and $h_a$ are some bandwidth parameters. The tuning parameters $h_x$ and $h_a$ are chosen according to the numerical results in Section 5 of Zhu et al. (2020a).

To evaluate the value function under the proposed I2DR $\tilde{d}^*(\cdot)$, we set $\pi^*$ to be a uniform density function and compute $\tilde{V}^*(\tilde{d})$, defined as

$$
\int x \int \frac{1}{|\tilde{d}(x)|} \left\{ \sum_{i \in G_{test}} Y_i K(h_x^{-1}(x - X_i))K(h_a^{-1}(\tilde{d}(x) - A_i)) \right\} \left\{ \sum_{i \in G_{test}} \frac{K(h_x^{-1}(x - X_i))}{|G_{test}|h_x^p} \right\} d\tilde{d} dx.
$$

The integration is calculated via a midpoint rule with a uniform grid.

Over 50 iterations, the average value functions of our proposed I2DRs computed by L-JIL and D-JIL are $-0.332$ and $-0.331$, larger than the value $-0.344$ of the IDR obtained by K-O-L. In the following, we focus on one particular iteration in L-JIL and D-JIL, respectively, to further interpret our results.

Finally, we show the results of one particular iteration under L-JIL whose value achieves the largest as $-0.307$, to illustrate the interpretability of the resulting I2DR. Under this iteration, L-JIL partitions $[0, 1]$ into three subintervals: $[0, 0.05)$, $[0.05, 0.17)$ and $[0.17, 1]$. Let $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ denote the corresponding regression coefficients associated with these three subintervals. We report these estimators in Table 2.5. According to Table 2.5, the proposed I2DR under L-JIL gives us a clear interpretation about the effect of baseline information.
on the dose assignment rule. For instance, patients whose genotype VKORC1 is AG or AA are more likely to receive low doses of Warfarin to prevent bleeding; older patients with larger weight shall be treated with higher dose levels. Future experiments are warranted to confirm these scientific findings. In addition, we also explore one particular iteration under D-JIL whose value nearly achieves the largest as \(-0.315\). Under this iteration, the I2DR under D-JIL recommends most patients to receive dose level in \([0, 0.05)\), \([0.15, 0.20)\) and \([0.20, 1]\), respectively. This finding accords with the partition results under L-JIL.

2.7 Discussions

2.7.1 Diverging Number of Change Points

Under Model I, we assume \(|\mathcal{P}_0|\) is fixed to simplify the results in Theorems 1, 2, 3, and 4. Our theoretical results can be generalized to the situation where \(|\mathcal{P}_0|\) diverges with \(n\) as well. Take L-JIL as an example. Similar to Theorem 6, we can show that the \(\ell_2\) integrated loss satisfies

\[
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|^2_2 da = O_p(|\mathcal{P}_0| n^{-1} \log n).
\]

Compared to the results in Theorem 1, the convergence rate here is slower by a factor \(|\mathcal{P}_0|\). In addition, \(|\mathcal{P}_0| = o(n/\log n)\) is required to guarantee the consistency of \(b_{\theta}(\cdot)\).

We next present more technical details. In the proof of Theorem 6 (see Section A.2.13 for details), we consider a more general framework and establish the \(\ell_2\) integrated loss of \(b_{\theta}(\cdot)\) by assuming \(\theta_0\) satisfies \(\limsup_{k \to \infty} k^{a_0} \text{AE}_k(\theta_0) < \infty\) where

\[
\text{AE}_k(\theta_0) = \inf_{\substack{\mathcal{P} : |\mathcal{P}| \leq k + 1, \theta \in \mathbb{R}^{|\mathcal{P}| + 1}}} \left\{ \sup_{a \in [0,1]} \left\| \theta_0(a) - \sum_{\mathcal{I} \in \mathcal{P}} \theta_0(\mathcal{I}) (a \in \mathcal{I}) \right\|_2 \right\}.
\]

By definition, \(\text{AE}_k(\theta_0)\) describes how well \(\theta_0(\cdot)\) can be approximated by a step function.

When \(\theta_0(\cdot)\) is a step function with number of jumps equal to \(|\mathcal{P}_0|\), we have \(\text{AE}_k(\theta_0) = 0\) for any \(k \geq |\mathcal{P}_0|\). As a result, \(\theta_0\) satisfies the condition \(\limsup_{k \to \infty} k^{a_0} \text{AE}_k(\theta_0) < \infty\) for any \(a_0 > 0\). As a result, the assertion (A.133) in the proof of Theorem 6 also holds for \(\theta_0(\cdot)\) and we have with probability at least \(1 - O(n^{-2})\) that

\[
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|^2_2 da \leq O(1)(|\mathcal{P}_0|^{-a_0} + \gamma_n |\mathcal{P}_0|),
\]

where \(O(1)\) denotes some positive constant. As \(|\mathcal{P}_0| \to \infty\) and \(a_0\) can be made arbitrarily
large, we have with probability at least $1 - O(n^{-2})$ that

$$
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2 da \leq O(1)(\gamma_n|\mathcal{P}_0|),
$$

where $O(1)$ denotes some positive constant.

In Theorem 6, we require $\gamma_n \gg n^{-1}\log n$. However, this condition can be relaxed to $\gamma_n \geq M_0 n^{-1}\log n$ for some sufficiently large constant $M_0 > 0$. Under the latter condition, we have

$$
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2 da = O_p(|\mathcal{P}_0|n^{-1}\log n).
$$

This yields the convergence rate of the $\ell_2$ integrated loss of $\hat{\theta}(\cdot)$.

### 2.7.2 Potential Alternative Approaches

In this chapter, we focus on modeling the outcome regression function to derive I2DR. Below, we outline two other potential approaches and discuss their weaknesses.

**A-learning Type Methods**

Let’s assume $q_{f}(\cdot)$ satisfies (2.1) and the partition $\mathcal{P}_0$ is known to us. In order to eliminate the baseline function $u_0(\cdot)$, we can apply Robinson’s transformation (see for example, Robinson 1988; Zhao et al. 2017; Chernozhukov et al. 2018, and the references therein) and compute $\tilde{q}_{f}$ by minimizing

$$
\arg \min_{\{q_{f} \in \mathcal{L} : \mathcal{P} \in \mathcal{P}_0\}} \frac{1}{n} \sum_{i=1}^{n} \left[ Y_i - \tilde{\mu}(X_i) - \sum_{\mathcal{I} \in \mathcal{P}_0} \{I(A_i \in \mathcal{I}) - \tilde{e}(\mathcal{I} | X_i)\} q_{f}(X) \right]^2,
$$

where $\tilde{\mu}(x)$ correspond to some nonparametric estimators for $\mathbb{E}(Y \mid X = x)$. Both $\tilde{\mu}$ and $\tilde{e}$ can be obtained by some generic machine learning methods with good prediction performance.

When $\mathcal{P}_0$ is unknown, one might consider estimating $\mathcal{P}_0$ and $\{q_{f} : \mathcal{I} \in \mathcal{P}_0\}$ jointly by

$$
\arg \min_{\{q_{f} \in \mathcal{L} : \mathcal{P} \in \mathcal{P}_0\}} \frac{1}{n} \sum_{i=1}^{n} \left[ Y_i - \tilde{\mu}(X_i) - \sum_{\mathcal{I} \in \mathcal{P}_0} \{I(A_i \in \mathcal{I}) - \tilde{e}(\mathcal{I} | X_i)\} q_{f}(X) \right]^2 + \gamma_n |\mathcal{P}|,
$$

for some tuning parameter $\gamma_n$. However, different from the objective function in (2.3),
for a given partition $\mathcal{P}$, all the functions $\{q_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}$ need to be jointed estimated. As a result, standard change point detection algorithms such as dynamic programming or binary segmentation (Scott and Knott 1974) cannot be applied. Exhaustive search among all possible partitions is computationally infeasible. It remains unknown how to efficiently solve the above optimization problem. We leave it for future research.

**Policy Search**

As commented in Section 2.2.2, to apply value search, we need to specify a preference function $\pi^\ast$. To better illustrate the idea, let us suppose $\pi^\ast(\cdot; x, \mathcal{I}) = p(a|x)/\int_{x' \in \mathcal{I}} p(a|x')dx'$. That is, the preference function is the same as the one we observe in our data. Then, for a given I2DR $d$, we can consider the following inverse propensity score weighted estimator for $V_{\pi^\ast}(d)$,

$$\hat{V}_{\pi^\ast}(d) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}(A_i \in d(X_i))}{\hat{e}(d(X_i)|X_i)} Y_i.$$

For a given partition $\mathcal{P}$, let $\mathcal{D}_\mathcal{P}$ denote the space of I2DRs that we consider. Then $\hat{d}$ can be computed by maximizing

$$\arg\max_{\mathcal{P} \in \mathcal{B}(m)} \arg\max_{d \in \mathcal{D}_\mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}(A_i \in d(X_i))}{\hat{e}(d(X_i)|X_i)} Y_i.$$

Suppose we consider the class of linear decision rules, i.e.,

$$\mathcal{D}_\mathcal{P} = \{d : d(x) = \arg\max_{\mathcal{I} \in \mathcal{P}} \theta^\top_{\mathcal{I}} \bar{x}\}.$$

It suffices to maximize

$$\arg\max_{\mathcal{P} \in \mathcal{B}(m)} \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{I} \in \mathcal{P}} \frac{\mathbb{I}(A_i \in \mathcal{I})}{\hat{e}(d(X_i)|X_i)} Y_i \{d(X_i) = \arg\max_{\mathcal{I} \in \mathcal{P}} q_\mathcal{I}(X_i)\}.$$

Similar to Section 2.7.2, for a given partition $\mathcal{P}$, all the functions $\{q_\mathcal{I} : \mathcal{I} \in \mathcal{P}\}$ need to be jointed estimated. As a result, dynamic programming cannot be applied. It remains unknown how to efficiently solve the above optimization problem. We leave it for future research.
2.7.3 Other Penalty Functions in L-JIL

In L-JIL, we use a ridge penalty in (2.6) to prevent overfitting in large $p$ problems. When the true regression coefficient $\theta_0(\cdot)$ is sufficiently sparse, one can consider replacing the ridge penalty with the LASSO Tibshirani (1996) to improve the estimation accuracy. However, optimizing the resulting objective function requires to compute the LASSO estimator $m(m - 1)/2$ times. This is far more computationally expensive than the proposed method. It remains unknown whether the computation can be simplified. We leave it for future research.
CHAPTER

3

DEEP JUMP LEARNING FOR OFF-POLICY EVALUATION IN CONTINUOUS TREATMENT SETTINGS

3.1 Introduction

As we discussed in Chapter 1, most existing works on off-policy evaluation (OPE) focus on discrete treatment settings. To handle continuous treatments, we develop a novel estimation method for OPE using deep jump learning. The key ingredient of our method lies in adaptively discretizing the treatment space using deep discretization, by leveraging deep learning and multi-scale change point detection. This allows us to apply existing OPE methods in discrete treatments to handle continuous treatments.

Our contributions are summarized as follows. Methodologically, we develop a deep jump learning (DJL) method by integrating deep learning (LeCun et al. 2015), multi-scale change point detection (see e.g., Niu et al. 2016, for an overview), and the doubly-robust value estimators in discrete domains. Our method does not require kernel bandwidth selection.
It does not suffer from the limitations of kernel-based methods. The key ingredient of our method lies in adaptively discretizing the treatment space using deep discretization. This allows us to apply the IPW or DR methods to derive the value estimator. The discretization addresses the first limitation of kernel-based methods, allowing us to handle discontinuous value functions. The adaptivity addresses the second limitation of kernel-based methods. Specifically, it guarantees the optimality of the proposed method in cases where the second-order derivative of the conditional mean function has an abrupt change in the treatment space. Theoretically, we derive the convergence rate of the value estimator under DJL for any policy of interest, allowing the conditional mean outcome to be either a continuous or piecewise function of the treatment; see Theorems 3.4.1 and 3.4.2 for details. Under the piecewise model assumption, the rate of convergence is faster than kernel-based OPE methods. Under the continuous model assumption, kernel-based estimators converge at a slower rate when the bandwidth undersmoothes or oversmoothes the data. Proofs of our theorems rely on establishing the uniform rate of convergence of deep learning estimators; see Lemma B.4.1 in the supplementary article. We expect this result to be of general interest in contributing to the line of work on developing theories for deep learning methods (see e.g., Imaizumi and Fukumizu 2019; Schmidt-Hieber et al. 2020; Farrell et al. 2021). Empirically, we show the proposed deep jump learning outperforms existing state-of-the-art OPE methods in both simulations and a real data application to warfarin dosing.

3.2 Preliminaries

We first formulate the OPE problem in continuous domains. We next review some related literature on the DR value estimator, kernel based evaluation methods, and multi-scale change point detection.

3.2.1 Problem Formulation

The observed offline datasets can be summarized into \{(X_i, A_i, Y_i)\}_{1 \leq i \leq n} where \(O_i = (X_i, A_i, Y_i)\) denotes the feature-treatment-outcome triplet for the \(i\)th individual and \(n\) denotes the total sample size. We assume these data triplets are independent copies of the population variables \((X, A, Y)\). Let \(\mathcal{X} \in \mathbb{R}^P\) and \(\mathcal{A}\) denote the \(p\) dimensional feature and treatment (or action) space, respectively. We focus on the setting where \(\mathcal{A}\) is one-dimensional, as in personalized dose finding and dynamic pricing. A decision rule or policy \(d : \mathcal{X} \rightarrow \mathcal{A}\) determines the treatment to be assigned given the observed feature. We use \(b\) to denote
the propensity score, also known as the behavior policy, that generates the observed data. Specifically, \( p(\bullet|X = x) \) denotes the probability density function of \( A \) given \( X = x \). Define the expected outcome function conditional on the feature-treatment pair as

\[
Q(x, a) = E(Y|X = x, A = a).
\]

As standard in the OPE and the causal inference literature (see e.g., Chen et al. 2016; Kallus and Zhou 2018), we assume the stable unit treatment value assumption, no unmeasured confounders assumption, and the positivity assumption are satisfied. The positivity assumption requires \( b \) to be uniformly bounded away from zero. The latter two assumptions are automatically satisfied in randomized studies. These three assumptions guarantee that a policy’s value is estimable from the observed data. Specifically, for a target policy \( d \) of interest, its value can be represented by

\[
V(d) = E[Q(X, d(X))].
\]

Our goal is to estimate the value \( V(d) \) based on the observed data.

### 3.2.2 Doubly Robust Estimator and Kernel-Based Evaluation

For discrete treatments, Dudík et al. (2011) proposed a DR estimator of \( V(d) \) by

\[
\frac{1}{n} \sum_{i=1}^{n} \psi(O_i, d, \hat{Q}, \hat{p}) = \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{Q}(X_i, d(X_i)) + \frac{I[A_i = d(X_i)]}{\hat{p}(A_i|X_i)} \{Y_i - \hat{Q}(X_i, A_i)\} \right],
\]

(3.1)

where \( I(\bullet) \) denotes the indicator function, \( \hat{Q} \) and \( \hat{p}(a|x) \) denote some estimators for the conditional mean function \( Q \) and the propensity score \( b \), respectively. The second term inside the bracket corresponds to an augmentation term. Its expectation equals zero when \( \hat{Q} = Q \). The purpose of adding this term is to offer additional protection against potential model misspecification of \( Q \). Such an estimator is doubly-robust in the sense that its consistency relies on either the estimation model of \( Q \) or \( p \) to be correctly specified. It can be semi-parametrically efficient whereas the importance sampling or direct method are not. By setting \( \hat{Q} = 0 \), (3.1) reduces to the IPW estimator.

In continuous treatment domains, the indicator function \( I[A_i = d(X_i)] \) equals zero almost surely. Consequently, naively applying (3.1) yields a plug-in estimator \( \sum_{i=1}^{n} \hat{Q}(X_i, d(X_i))/n \). To address this concern, the kernel-based methods proposed to replace the indicator func-
tion in (3.1) with a kernel function \( K[\{A_i - d(X_i)\}/h] \), i.e.,

\[
\frac{1}{n} \sum_{i=1}^{n} \psi_i(O_i, d_i, \hat{Q}, \hat{p}) = \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{Q}(X_i, d(X_i)) + \frac{K\{\frac{A_i - d(X_i)}{h}\}}{\hat{p}(A_i|X_i)} \{Y_i - \hat{Q}(X_i, A_i)\} \right].
\] (3.2)

Here, the bandwidth \( h \) represents a trade-off. The variance of the resulting value estimator decays with \( h \). Yet, its bias increases with \( h \). More specifically, it follows from Theorem 1 of Kallus and Zhou (2018) that the leading term of the bias is equal to

\[
h^2 \int \frac{u^2 K(u) du}{2} \mathbb{E}\left\{ \frac{\partial^2 Q(X, a)}{\partial a^2} \bigg|_{a=d(X)} \right\}. \] (3.3)

To ensure the term in (3.3) decays to zero as \( h \) goes to 0, it requires the expected second derivative of the function \( Q(x, a) \) exists, and thus \( Q(x, a) \) needs to be a smooth function of \( a \). However, as commented in the introduction, this assumption could be violated in certain applications.

### 3.2.3 Multi-Scale Change Point Detection

To adaptively discretize the treatment space, we leverage ideas from multi-scale change point detection literature. The change point analysis considers an ordered sequence of data, \( Y_{i:n} = \{Y_1, \ldots, Y_n\} \), with unknown change point locations, \( \tau = \{\tau_1, \ldots, \tau_K\} \) for some unknown integer \( K \). Here, \( \tau_i \) is an integer between 1 and \( n - 1 \) inclusive, and satisfies \( \tau_i < \tau_j \) for \( i < j \). These change points split the data into \( K + 1 \) segments. Assume there are sufficiently many data points lying within each segment such that the expected reward can be consistently estimated. Within each segment, the expected outcome is a constant function; see the left panel of Figure 3.1 for details. A number of methods have been proposed on estimating change points (see for example, Boysen et al. 2009; Killick et al. 2012b; Frick et al. 2014; Fryzlewicz 2014, 2020, and the references therein), by minimizing a penalized objective function:

\[
\arg\min_{\tau} \left( \frac{1}{n} \sum_{i=1}^{K+1} [\mathcal{C}\{Y_{(\tau_{i-1}+1):\tau_i}\} + \gamma_n K] \right),
\]

where \( \mathcal{C} \) is a cost function that measures the goodness-of-the-fit of the constant function within each segment and \( \gamma_n K \) penalizes the number of change points. We remark that all the above cited works focused on either models without features or linear models. Our proposal goes beyond these works in that we consider models with features and use deep
neural networks (DNN) to capture the complex relationship between the outcome and features.

3.3 Deep Jump Learning

In Section 3.3.1, we use a toy example to demonstrate the limitation of kernel-based methods. We present the main idea of our algorithm in Section 3.3.2. Details are given in Section 3.3.3. For simplicity, we set the action space \( \mathcal{A} = [0, 1] \). Define a discretization \( \mathcal{D} \) for the treatment space \( \mathcal{A} \) as a set of mutually disjoint intervals \( \{[\tau_0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_{K-1}, \tau_K]\} \) for some \( 0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{K-1} < \tau_K = 1 \) and some integer \( K \geq 1 \). The union of these intervals covers \( \mathcal{A} \). We use \( J(\mathcal{D}) \) to denote the set of change point locations, i.e., \( \{\tau_1, \cdots, \tau_{K-1}\} \). We use \( |\mathcal{D}| \) to denote the number of intervals in \( \mathcal{D} \) and \( |\mathcal{I}| \) to denote the length of any interval \( \mathcal{I} \).

3.3.1 Toy Example

As discussed in the introduction, existing kernel-based methods use a single bandwidth to construct the value estimator. Ideally, the bandwidth \( h \) in the kernel \( K \left( \frac{A_i - d(X_i)}{h} \right) \) shall vary with \( d(X_i) \) to improve the accuracy of the value estimator. To elaborate this, consider the function \( Q(x, a) = 10 \max(a^2-0.25, 0) \log(x+2) \) for any \( x, a \in [0, 1] \). By definition, \( Q(x, a) \) is smooth over the entire feature-treatment space. However, it has different patterns when the treatment belongs to different intervals. Specifically, for \( a \in [0, 0.5] \), \( Q(x, a) \) is constant as a function of \( a \). For \( a \in (0.5, 1] \), \( Q(x, a) \) depends quadratically in \( a \). See the middle panel of Figure 3.1 for details.

Consider the target policy \( d(x) = x \). We decompose the value \( V(d) \) into \( V^{(1)}(d) + V^{(2)}(d) \) where

\[
V^{(1)}(d) = \mathbb{E}[Q\{X, d(X)\}I\{d(X) \leq 0.5\}], \text{ and } V^{(2)}(d) = \mathbb{E}[Q\{X, d(X)\}I\{d(X) > 0.5\}].
\]

Similarly, denote the corresponding kernel-based value estimators by

\[
\hat{V}^{(1)}(d; h) = \frac{1}{n} \sum_{i=1}^{n} [\psi_h I\{d(X_i) \leq 0.5\}], \text{ and } \hat{V}^{(2)}(d; h) = \frac{1}{n} \sum_{i=1}^{n} [\psi_h I\{d(X_i) > 0.5\}]
\]

where \( \psi_h := \psi_h(O_i, d, \hat{Q}, \hat{\mathcal{D}}) \) is defined in (3.2). Since \( Q(x, a) \) is a constant function of \( a \in [0, 0.5] \), its second-order derivative \( \partial^2 Q(x, a)/\partial a^2 \) equals zero. In view of (3.3), when \( d(x) \leq 0.5 \)
Figure 3.1: Left panel: example of piece-wise constant function. Middle panel: the oracle conditional mean function $Q$ on the feature-treatment space for the toy example. Right panel: the green curve presents the oracle $Q\{x, d(x)\}$ under target policy $d(x) = x$ in the toy example; and the red curve is the fitted mean value by DJL and the pink dash line corresponds to the 95% confidence bound.

Table 3.1: The absolute error and the standard deviation (in parentheses) of the estimated values for $V^{(1)}$ and $V^{(2)}$, using DJL and kernel-based methods, for target policy $d(x) = x$ in the toy example.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Indicator</th>
<th>Deep Jump Learning</th>
<th>Kernel with $h = 0.4$</th>
<th>Kernel with $h = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V^{(1)}(d)$</td>
<td>$I{d(X) \leq 0.5}$</td>
<td>0.31 (0.06)</td>
<td>0.50 (0.08)</td>
<td>0.40 (0.05)</td>
</tr>
<tr>
<td>$V^{(2)}(d)$</td>
<td>$I{d(X) &gt; 0.5}$</td>
<td>0.09 (0.19)</td>
<td>0.16 (0.20)</td>
<td>1.09 (0.09)</td>
</tr>
</tbody>
</table>

0.5, the bias of $\hat{V}^{(1)}(d; h)$ will be small even with a sufficiently large $h$. As such, a large $h$ is preferred to reduce the variance of $\hat{V}^{(1)}(d; h)$. When $d(x) > 0.5$, a small $h$ is preferred to reduce the bias of $\hat{V}^{(2)}(d; h)$. A simulation study is provided to demonstrate the drawback of the kernel-based methods. Specifically, we set $X, A \sim \text{Uniform}[0, 1]$ and generate $Y|X, A \sim N\{Q(X, A), 1\}$. We apply the kernel-based methods with a Gaussian kernel to estimate $V^{(1)}(d)$ and $V^{(2)}(d)$ with the sample size $n = 300$ over 100 replications. See Table 3.1 for details of the absolute error and standard deviation of $\hat{V}^{(1)}(d; h)$ and $\hat{V}^{(2)}(d; h)$ with two different bandwidths $h = 0.4$ and 1. It can be seen that due to the use of a single bandwidth, the kernel-based estimator suffers from either a large absolute error or a large variance.

To overcome this limitation, we propose to adaptively discretize the treatment space into a union of disjoint intervals such that within each interval $\mathcal{I}$, the conditional mean function $Q$ can be well-approximated by some functions $q_{\mathcal{I}}$ that depend on features but not on the treatment (constant in $a$), i.e., $Q(\bullet, a) \approx \sum_{\mathcal{I} \in \mathcal{D}} \left[ I[a \in \mathcal{I}] q_{\mathcal{I}}(\bullet) \right]$. By the discretization, one can apply the IPW or DR methods to evaluate the value. The advantage of adaptive discretization is illustrated in the right panel of Figure 3.1, where we apply the proposed DJL method to the toy example. See details of the proposed method and its implementation in
Sections 3.3.2 and 3.3.3. When \( a \leq 0.5 \), \( Q(x, a) \) is constant in \( a \). It is likely that our procedure will not further split the interval \([0, 0.5]\). Consequently, the corresponding DR estimator for \( V^{(1)}(d) \) will not suffer from large variance. When \( a > 0.5 \), our procedure will split \((0.5, 1]\) into a series of sub-intervals, approximating \( Q \) by a step function. This guarantees the resulting DR estimator for \( V^{(2)}(d) \) will not suffer from large bias. Consequently, the proposed value estimator achieves a smaller mean squared error than kernel-based estimators. See Table 3.1 for details.

### 3.3.2 The Main Idea

We consider the following two model assumptions, which cover a variety of scenarios in practice.

**Model 1: Piecewise function.** Suppose

\[
Q(x, a) = \sum_{\mathcal{I} \in D_0} \left\{ q_{\mathcal{I}0}(x)\mathbb{I}(a \in \mathcal{I}) \right\}, \quad \text{for any } x \in \mathcal{X}, \text{ for any } a \in \mathcal{A},
\]

for some partition \( D_0 \) of \([0, 1]\) and a collection of functions \( \{ q_{\mathcal{I}0} \}_{\mathcal{I} \in D_0} \).

**Model 2: Continuous function.** Suppose \( Q \) is a continuous function of \( a \) and \( x \).

Model 1 covers the dynamic pricing example we mentioned in the introduction. In our simulation studies in Section 3.5.1, the underlying model is set to be a piecewise function in Scenarios 1 and 2. Model 2 covers the personalized dose-finding example, Scenarios 3 and 4 in our simulation studies, as well as the real data section in Section 3.5.2. We next detail the proposed method, which will work when either Model 1 or 2 holds.

Motivated by Model 1, our goal is to identify an optimal discretization \( \widehat{\mathcal{D}} \) such that for each interval \( \mathcal{I} \in \widehat{\mathcal{D}} \), \( Q(x, a) \) is approximately a constant function of \( a \in \mathcal{I} \). Specifically, under Model 1, we assume the function \( Q(x, a) \) is a piecewise function on the action space. Within each segment \( \mathcal{I} \), the function \( Q(x, a) \) is a constant function of \( a \), but can be any function of the features \( x \). In other words, \( Q(x, a_1) = Q(x, a_2) \) for any \( a_1, a_2 \in \mathcal{I} \). Thus, we denote the function \( Q(x, a) \) at each segment \( \mathcal{I} \) as \( q_{\mathcal{I}}(x) \), which yields a piecewise function \( Q(x, a) = \sum_{\mathcal{I}} q_{\mathcal{I}}(x)\mathbb{I}(a \in \mathcal{I}) \), as stated in (3.4). In the real applications, the true function \( Q(x, a) \) could be either a continuous function, or a piecewise function. As such, we propose to approximate the underlying unknown function \( Q(x, a) \) by these piecewise functions of \( a \) using the proposed DJL method. Such approximation allows us to derive the DR estimator based on \( \widehat{\mathcal{D}} \). The bias and variance of the resulting estimator are largely affected by the number of intervals in \( \widehat{\mathcal{D}} \). Specifically, if \( |\widehat{\mathcal{D}}| \) is too small, then the piecewise approximation
is not accurate, leading to a biased estimator. If $|\mathcal{D}|$ is too large, then $\mathcal{D}$ will contain many short intervals, and the resulting estimator might suffer from a large variance.

To this end, we develop a data-adaptive method to compute $\mathcal{D}$. We first divide the treatment space $\mathcal{A}$ into $m$ disjoint intervals: $[0, 1/m), [1/m, 2/m), \ldots, [(m-1)/m, 1]$. We require the integer $m$ to diverge with the sample size $n$, such that the conditional mean function $Q$ can be well-approximated by a piecewise function on these intervals. Note that these $m$ initial intervals is not equal to $\mathcal{D}$, but only serve as the initial candidate intervals. Yet, $\mathcal{D}$ will be constructed by adaptively combining some of these intervals. We find in our numerical studies that the size of the final partition $|\mathcal{D}|$ is usually much less than $m$ (see Table B.4 in Appendix B.2 for more details). In practice, we recommend to set the initial number of intervals $m$ to be proportional to the sample size $n$, i.e., $m = n/c$ for some constant $c > 0$. The performance of the resulting value estimator is not overly sensitive to the choice of $c$.

We define $\mathcal{B}(m)$ as the set of discretizations $\mathcal{D}$ such that each interval $\mathcal{I} \in \mathcal{D}$ corresponds to a union of some of the $m$ initial intervals. Each discretization $\mathcal{D} \in \mathcal{B}(m)$ is associated with a set of functions $\{q_{\mathcal{I}} : \mathcal{I} \in \mathcal{D}\}$. We model these $q_{\mathcal{I}}$ using DNNs, to capture the complex dependence between the outcome and features. When $Q(\cdot, a)$ is well-approximated by $\sum_{\mathcal{I} \in \mathcal{D}} [\mathbb{I}(a \in \mathcal{I}) q_{\mathcal{I}}(\cdot)]$, we expect the least square loss $\sum_{\mathcal{I} \in \mathcal{D}} \sum_{i=1}^{n} [\mathbb{I}(A_i \in \mathcal{I}) \{Y_i - q_{\mathcal{I}}(X_i)\}]^2$, will be small. Thus, $\mathcal{D}$ can be estimated by solving

$$\left(\hat{\mathcal{D}}, \left\{\hat{q}_{\mathcal{I}} : \mathcal{I} \in \hat{\mathcal{D}}\right\}\right) = \arg \min_{\mathcal{D} \in \mathcal{B}(m), \left\{q_{\mathcal{I}} \in \mathcal{Q}_{\mathcal{I}} : \mathcal{I} \in \mathcal{D}\right\}} \left(\sum_{\mathcal{I} \in \mathcal{D}} \left[\frac{1}{n} \sum_{i=1}^{n} [\mathbb{I}(A_i \in \mathcal{I}) \{Y_i - q_{\mathcal{I}}(X_i)\}]^2\right] + \gamma_n |\mathcal{D}|\right), \quad (3.5)$$

for some regularization parameter $\gamma_n$ and some function class of DNNs $\mathcal{Q}_{\mathcal{D}}$. Here, the penalty term $\gamma_n |\mathcal{D}|$ in (3.5) controls the total number of intervals in $\hat{\mathcal{D}}$, as in multi-scale change point detection. A large $\gamma_n$ results in few intervals in $\hat{\mathcal{D}}$ and a potential large bias of the value estimator, whereas a small $\gamma_n$ procedures a large number of intervals in $\hat{\mathcal{D}}$, leading to a noisy value estimator. The theoretical order of $\gamma_n$ is detailed in Section 3.4. In practice, we use cross-validation to select $\gamma_n$. We refer to this step as deep discretization. Details of solving (3.5) are given in Section 3.3.3.

Given $\hat{\mathcal{D}}$ and $\left\{\hat{q}_{\mathcal{I}} : \mathcal{I} \in \hat{\mathcal{D}}\right\}$, we apply the DR estimator in (3.1) to derive the value estimate for any target policy of interest $d$, i.e.,

$$\hat{V}^{DR}(d) = \frac{1}{n} \sum_{\mathcal{I} \in \hat{\mathcal{D}}} \sum_{i=1}^{n} \left[\mathbb{I}(d(X_i) \in \mathcal{I}) \left\{\mathbb{I}(A_i \in \mathcal{I}) \frac{\hat{q}_{\mathcal{I}}(X_i)}{\hat{p}_{\mathcal{I}}(X_i)} \{Y_i - \hat{q}_{\mathcal{I}}(X_i)\} + \hat{q}_{\mathcal{I}}(X_i)\right\}\right], \quad (3.6)$$
where $\hat{p}_x(x)$ is some estimator of the generalized propensity score function $\Pr(A \in \mathcal{I} | X = x)$. We call this method as the deep jump learning. We remark that the proposed method yields a consistent value estimator allowing the function $Q$ to be either a continuous or piecewise function of the treatment. Under different model assumptions, we derive the corresponding rate of convergence of our method in Section 3.4.

### 3.3.3 The Complete Algorithm for Deep Jump Learning

We present the details for DJL in this section. To further reduce the bias of the value estimator in (3.6), we employ a data splitting and cross-fitting strategy (Chernozhukov et al. 2017). That is, we use different subsets of data samples to estimate the discretization $\mathcal{D}$ and to construct the value estimator. Our algorithm consists of three steps: data splitting, deep discretization, and cross-fitting. We detail each of these steps below.

**Step 1: data splitting.** We divide all $n$ samples into $L$ disjoint subsets of equal size, where $L_\ell$ denotes the indices of samples in the $\ell$th subset for $\ell = 1, \cdots, L$. Let $L_\ell^c = \{1, 2, \cdots, n\} - L_\ell$ as the complement of $L_\ell$. Data splitting allows us to use one part of the data, i.e., $L_\ell^c$, to train machine learning models for the conditional mean function and propensity score function, and the remaining part, i.e., $L_\ell$, to estimate the value. We aggregate the resulting estimates over different subsets to get full efficiency, as summarized in the third step.

**Step 2: deep discretization.** For each $\ell = 1, \cdots, L$, we propose to apply deep discretization to compute a discretization $\mathcal{D}_\ell$ and $\{\tilde{q}_\mathcal{I}(\ell) : \mathcal{I} \in \mathcal{D}_\ell\}$ by solving a version of (3.5) using the data subset in $L_\ell^c$ only. We next present the computational details for solving this optimization. Our algorithm employs the pruned exact linear time method (Killick et al. 2012b) to identify the change points with a cost function that involves DNN training. Specifically, for any interval $\mathcal{I}$, define $\tilde{q}_\mathcal{I}(\ell)$ as the minimizer of

$$
\arg\min_{q_{\mathcal{I}} \in \mathbb{Q}_{\mathcal{I}}} \frac{1}{|L_\ell^c|} \sum_{i \in L_\ell^c} \left[ \mathbb{1}(A_i \in \mathcal{I}) \{ q_{\mathcal{I}}(X_i) - Y_i \}^2 \right],
$$

where $|L_\ell^c|$ denotes the number of samples in $L_\ell^c$. Define the cost function $\mathcal{C}(\mathcal{I})$ as the minimum value of the objective function (3.7), i.e,

$$
\mathcal{C}(\mathcal{I}) = \frac{1}{|L_\ell^c|} \sum_{i \in L_\ell^c} \left[ \mathbb{1}(A_i \in \mathcal{I}) \{ \tilde{q}_\mathcal{I}(\ell)(X_i) - Y_i \}^2 \right].
$$

Computation of $\mathcal{D}_\ell$ relies on dynamic programming (Friedrich et al. 2008). For any integer
Global: data \( \{(X_i, A_i, Y_i)\}_{1 \leq i \leq n} \); number of initial intervals \( m \); penalty term \( \gamma_n \); target policy \( d \).

Local: Bellman function \( \text{Bell} \in \mathbb{R}^m \); partitions \( \mathcal{D} \); DNN functions \( \{\hat{q}_{\mathcal{S}}, \hat{p}_{\mathcal{S}} : \mathcal{I} \in \mathcal{D}\} \); a vector \( \tau \in \mathbb{N}^m \); a set of candidate point lists \( \mathcal{R} \).

Output: the value estimator for target policy \( \hat{V}(d) \).

I. Split all \( n \) samples into \( \mathcal{L} \) subsets as \( \{\mathcal{L}_1, \cdots, \mathcal{L}_\mathcal{L}\} \); \( \hat{V}(d) \leftarrow 0 \);

II. Initialization:

1. Set even segment on the action space with \( m \) pieces:
   \( \mathcal{I} = \{(0, 1/m) m \}, (1/m, 2/m), \cdots, ((m-1)/m, 1) \};

2. Create a function to calculate cost \( \mathcal{C} \) with inputs \( (l, r) \):
   - (i). Let \( \mathcal{I} = [l/m, r/m) \) if \( r < m \) else \( \mathcal{I} = [l/m, 1] \);
   - (ii). Fit a DNN regressor: \( \hat{q}_{\mathcal{S}}(\cdot) \leftarrow \mathbb{I}(i \in \mathcal{L}_i)\mathbb{I}(A_i \in \mathcal{I})Y_i \sim \mathbb{I}(A_i \in \mathcal{I})\mathcal{DNN}(X_i) \);
   - (iii). Store the cost: \( \mathcal{C}(\mathcal{I}) \leftarrow \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{I}) \{\hat{q}_{\mathcal{S}}(X_i) - Y_i\}^2 \);
   - Return \( \mathcal{C}(l, r) \);

III. For \( l = 1, \cdots, \mathcal{L} \):

1. Set the training dataset as \( \mathcal{L} = \{1, 2, \cdots, n\} \) \( \mathcal{L}_i \);

2. \( \text{Bell}(0) \leftarrow -\gamma_n \); \( \mathcal{D} = [0, 1] \); \( \tau \leftarrow \text{Null} \); \( \mathcal{R}(0) \leftarrow \{0\} \);

3. Apply the pruned exact linear time method to get partitions: For \( v^* = 1, \cdots, m \):
   - (i). \( \text{Bell}(v^*) = \min_{v \in \mathcal{D}(v^*)} \{\text{Bell}(v) + \mathcal{C}(\{v/m, v^*/m\}) + \gamma_n\} \);
   - (ii). \( v^1 \leftarrow \arg\min_{v \in \mathcal{D}(v^*)} \{\text{Bell}(v) + \mathcal{C}(\{v/m, v^*/m\}) + \gamma_n\} \);
   - (iii). \( \tau(v^*) \leftarrow \{v^1, \tau(v^1)\} \);
   - (iv). \( \mathcal{R}(v^*) \leftarrow \{v \in \mathcal{R}(v^* - 1) \cup \{v^* - 1\} : \text{Bell}(v) + \mathcal{C}(\{v/m, (v^* - 1)/m\}) \leq \text{Bell}(v^* - 1)\} \);

4. Construct the DR value estimator: \( r \leftarrow m \); \( l \leftarrow \tau(r) \); While \( r > 0 \):
   - (i). \( \text{Bell}(v^*) = \min_{v \in \mathcal{D}(v^*)} \{\text{Bell}(v) + \mathcal{C}(\{v/m, v^*/m\}) + \gamma_n\} \);
   - (ii). \( v^1 \leftarrow \arg\min_{v \in \mathcal{D}(v^*)} \{\text{Bell}(v) + \mathcal{C}(\{v/m, v^*/m\}) + \gamma_n\} \);
   - (iii). \( \tau(v^*) \leftarrow \{v^1, \tau(v^1)\} \);
   - (iv). \( r \leftarrow l \); \( l \leftarrow \tau(r) \);

5. Evaluation using testing dataset \( \mathcal{L}_i \):
   \( \hat{V}(d) = \sum_{\mathcal{S} \in \mathcal{D}} \left(\sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{S}) \left[1_{d(X_i) \in \mathcal{S}}(\{Y_i - \hat{q}_{\mathcal{S}}(X_i)\} + \hat{q}_{\mathcal{S}}(X_i))\right]\right) \); \( \text{return} \hat{V}(d)/n \).

Algorithm 2: Deep Jump Learning

\( 1 \leq v^* < m \), denote by \( \mathcal{B}(m, v^*) \) the set consisting of all possible discretizations \( \mathcal{D}_{v^*} \) of \( [0, v^*/m) \). Set \( \mathcal{B}(m, m) = \mathcal{B}(m) \), we define the Bellman function as

\[
\text{Bell}(v^*) = \inf_{\mathcal{D}_{v^*} \in \mathcal{B}(m, v^*)} \left\{ \sum_{\mathcal{I} \in \mathcal{D}_{v^*}} \mathcal{C}(\mathcal{I})(\mathcal{I}) + \gamma_n(\{|\mathcal{D}_{v^*}| - 1\}) \right\}, \text{and Bell}(0) = -\gamma_n.
\]
Our algorithm recursively updates the Bellman cost function for \( v^* = 1, 2, \cdots \) by

\[
\text{Bell}(v^*) = \min_{v \in \mathcal{R}_{v^*}} \left\{ \text{Bell}(v) + \mathcal{G}^{(l)}([v/m, v^*/m]) + \gamma_n \right\}, \quad \text{for any } v^* \geq 1, \tag{3.8}
\]

where \( \mathcal{R}_{v^*} \) is the candidate change points list. For a given \( v \), the right-hand-side of (3.8) corresponds to the cost of partitioning on a particular point. We then identify the best \( v \) that minimizes the cost. This yields the Bellman function on \([0, v^*/m]\) on the left-hand-side. In other words, (3.8) is a recursive formula used in our dynamic algorithm to update the Bellman equation for the locations of change points. It is recursive as the Bellman function appears on both sides of (3.8). Here, the list of candidate change points \( \mathcal{R}_{v^*} \) is given by

\[
\left\{ v \in \mathcal{R}_{v^*-1} \cup \{ v^*-1 \} : \text{Bell}(v) + \mathcal{G}^{(l)}([v/m, (v^*-1)/m]) \leq \text{Bell}(v^*-1) \right\}, \tag{3.9}
\]

during each iteration with \( \mathcal{R}_0 = \{0\} \). The constraint listed in (3.9) restricts the research space in (3.8) to a potentially much smaller set of candidate change points, i.e., \( \mathcal{R}_{v^*} \). The main purpose is to facilitate the computation by discarding change points not relevant to obtain the final discretization. It yields a linear computational cost (Killick et al. 2012b). In contrast, without these restrictions, it would yield a quadratic computational cost (Friedrich et al. 2008).

To solve (3.8), we search the optimal change point location \( v \) that minimizes \( \text{Bell}(v^*) \). This requires deep learning to estimate \( \tilde{\mathcal{G}}^{(l)} \) and \( \mathcal{G}^{(l)}(\mathcal{I}) \) with \( \mathcal{I} = [v/m, v^*/m] \) for each \( v \in \mathcal{R}_{v^*} \). Let \( v^1 \) be the corresponding minimizer. We then define the change points list \( \tau(v^*) \) as the set of change-point locations in \([0, v^*/m]\) computed by the dynamic programming algorithm. It is computed iteratively based on the update \( \tau(v^*) = \{ v^1, \tau(v^1) \} \), which means that during each iteration, it includes the current best change point location \( v^1 \) (that minimizes (3.8)) and the previous change-point list for the interval \([0, v^1/m]\). This procedure is iterated to compute \( \text{Bell}(v^*) \) and \( \tau(v^*) \) for \( v^* = 1, \cdots, m \), to find the best change-point set for interval \([0, 1]\). The optimal partition \( \widehat{\mathcal{D}}^{(l)} \) is determined by the values stored in \( \tau \). Specifically, we initialize \( \widehat{\mathcal{D}}^{(l)} = [\tau(m)/m, 1] \), \( r = m \) and recursively update \( \widehat{\mathcal{D}}^{(l)} \) by setting \( \widehat{\mathcal{D}}^{(l)} \leftarrow \widehat{\mathcal{D}}^{(l)} \cup [\tau(r)/m, r/m] \) and \( r \leftarrow \tau(r) \), as in dynamic programming (Friedrich et al. 2008).

**Step 3: cross-fitting.** For each interval in the estimated optimal partition \( \widehat{\mathcal{D}}^{(l)} \), let \( \widehat{p}_{\mathcal{I}}^{(l)}(x) \) denote some estimator for the propensity score \( \Pr(A \in \mathcal{I} | X = x) \). In a randomized study, the density function \( p(a|x) \) is known to us and we set \( \widehat{p}_{\mathcal{I}}^{(l)}(x) = \int_{a \in \mathcal{I}} p(a|x) da \). To deal with data from observational studies, we estimate the generalized propensity score with deep learning using the training dataset \( \mathbb{L}_I \) as \( \widehat{p}_{\mathcal{I}}^{(l)}(x) \). We evaluate the target policy in each
subsample \( L_{\ell} \), based on the estimators \((\hat{q}^{(\ell)}_{y}, \hat{p}^{(\ell)}_{y}, \text{ and } \hat{\Theta}^{(\ell)})\) trained in its complementary subsamples \( L_{c\ell}^{c} = \{1, \cdots, n\} - L_{\ell} \). Denote this value estimator for subset \( L_{\ell} \) as \( \hat{V}_{\ell} \). The final proposed value estimator for \( V(d) \) is to aggregate over \( \hat{V}_{\ell} \) for \( \ell = 1, \cdots, \mathcal{L} \) via cross-fitting, given by,

\[
\hat{V}(d) = \frac{1}{n} \sum_{\ell=1}^{\mathcal{L}} \sum_{I \in \mathcal{J}} \sum_{i \in L_{\ell}} \left( \mathbb{I}(A_{i} \in \mathcal{J}) \frac{I(d(X_{i}) \in \mathcal{J})}{\hat{p}^{(\ell)}_{y}(X_{i})} \{ Y_{i} - \hat{q}^{(\ell)}_{y}(X_{i}) \} + \mathbb{I}(A_{i} \in \mathcal{J}) \hat{q}^{(\ell)}_{y}(X_{i}) \right).
\]  

(3.10)

Note the samples used to construct \( \hat{V} \) inside bracket are independent from those to estimate \( \hat{q}^{(\ell)}_{y} \), \( \hat{p}^{(\ell)}_{y} \) and \( \hat{\Theta}^{(\ell)} \). This helps remove the bias induced by overfitting in the estimation of \( \hat{q}^{(\ell)}_{y} \), \( \hat{p}^{(\ell)}_{y} \) and \( \hat{\Theta}^{(\ell)} \).

We give the full detailed pseudocode in Algorithm 2. The computational complexity required to implement the proposed approach is \( O(mCn) \), where \( Cn \) is the computational complexity of training one DNN model with the sample size \( n \). Detailed analysis is provided in Section B.1 in Appendix. The code is publicly available at our repository at https://github.com/HengruiCai/DJL.

### 3.4 Theory

We investigate the theoretical properties of the proposed DJL method. All the proofs are provided in the supplementary article. Without loss of generality, assume the support \( \mathcal{X} = [0, 1]^p \). To simplify the analysis, we focus on the case where the behavior policy \( b \) is known to us, which automatically holds for data from randomized studies. We focus on the setting where the conditional mean function \( Q \) is a smooth function of the features; see A1 below. Specifically, define the class of \( \beta \)-smooth functions, also known as Hölder smooth functions with exponent \( \beta \), as

\[
\Phi(\beta, c) = \left\{ h : \sup_{\|a\|_1 \leq \|\beta\|} \sup_{x \in \mathcal{X}} |\Delta^ah(x)| \leq c, \sup_{\|a\|_1 = \|\beta\|} \sup_{x, z \in \mathcal{X}, x \neq z} \frac{|\Delta^ah(x) - \Delta^ah(z)|}{\|x - z\|^{\beta - \|\beta\|}_2} \leq c \right\},
\]

for some constant \( c > 0 \), where \( \|\beta\| \) denotes the largest integer that is smaller than \( \beta \) and \( \Delta^a \) denotes the differential operator \( \Delta^a \) denote the differential operator: \( \Delta^ah(x) = \partial^{\|a\|}_a h(x)/\partial x_1^{a_1} \cdots \partial x_p^{a_p} \), where \( x = [x_1, \ldots, x_p] \). When \( \beta \) is an integer, \( \beta \)-smoothness essentially requires a function to have bounded derivatives up to the \( \beta \)th order. The Hölder smoothness assumption is commonly imposed in the current literature (see e.g., Farrell
et al. 2021), which is a special example of the function classes that can be learned by neural nets. Meanwhile, the proposed DJL method is valid when $Q(x, a)$ is a nonsmooth function of $x$ as well (see e.g., Imaizumi and Fukumizu 2019). Our theory thus can be further generalized to any function class that can be learned by neural nets at a certain rate. We introduce the following conditions.

(A1.) Suppose $p(a|\cdot) \in \Phi(\beta, c)$, and $Q(\cdot, a) \in \Phi(\beta, c)$ for any $a$.

(A2.) Functions $\{b_{q_I}: I \in \mathcal{D}(\ell)\}$ are uniformly bounded.

Assumption (A2) ensures that the optimizer would not diverge in the uniform norm sense. Similar assumptions are commonly imposed in the literature to derive the convergence rates of DNN estimators (see e.g., Farrell et al. 2021). Combining (A2) with (A1) allows us to derive the uniform rate of convergence for the class of DNN estimators $\{b_{q_I}: I \in \mathcal{D}\}$.

Specifically, $b_{q_I}$ converges at a rate of $O_p\{n |I|^{-2\beta/(2\beta + p)}\}$ where the big-$O$ terms are uniform in $I$, $p$ is the dimension of features. See Lemma B.4.1 in the supplementary article for details.

3.4.1 Properties under Model 1

We first consider Model 1 where the function $Q(x, a)$ takes the form of (3.4). As commented, this assumption holds in applications such as dynamic pricing. Without loss of generality, assume $q_{\mathcal{S}_I,0} \neq q_{\mathcal{S}_J,0}$ for any two adjacent intervals $\mathcal{S}_I, \mathcal{S}_J \in \mathcal{D}_0$. This guarantees that the representation in (3.4) is unique. Let $L_\mathcal{S}$ and $W_\mathcal{S}$ be the number of hidden layers and total number of parameters in the function class of DNNs $\mathcal{Q}_\mathcal{S}$. Assume the number of change points in $\mathcal{D}_0$ is fixed. The following theorem summarizes the rate of convergence of the proposed estimator under Model 1.

Theorem 3.4.1. Suppose (3.4), (A1) and (A2) hold. Suppose $m$ is proportional to $n$, $Y$ is a bounded variable and $A$ has a bounded probability density function on $[0, 1]$. Assume $\{\gamma_n\}_{n \in \mathbb{N}}$ satisfies $\gamma_n \to 0$ and $\gamma_n \gg n^{-2\beta/(2\beta + p)} \log^8 n$. Then, there exist some classes of DNNs $\{\mathcal{Q}_\mathcal{S}: \mathcal{S} \in \mathcal{T}\}$ with $L_\mathcal{S} \asymp \log(n|\mathcal{S}|)$ and $W_\mathcal{S} \asymp (n|\mathcal{S}|)^{p/(2\beta + p)} \log(n|\mathcal{S}|)$ such that the following events occur with probability at least $1 - O(n^{-2})$,

(i) $|\mathcal{T}| = |\mathcal{D}_0|$; and (ii) $\max_{\tau \in J(\mathcal{D}_0)} \min_{\tau \in J(\mathcal{T})} |\hat{\tau} - \tau| = O\{n^{-2\beta/(2\beta + p)} \log^8 n\}$.

In addition, for any policy $d$ such that for any $\tau_0 \in J(\mathcal{D}_0)$, $\Pr(d(X) \in [\tau_0 - \epsilon, \tau_0 + \epsilon]) = O(\epsilon)$,

(iii) $V(d) = V(d) + O_p\{n^{-2\beta/(2\beta + p)} \log^8 n\} + O_p(n^{-1/2})$.

We make a few remarks. First, the result in (i) imply that deep discretization correctly identifies the number of change points. The result in (iii) imply that any change point in $\mathcal{D}_0$
can be consistently identified. In particular, \( J(D) \) corresponds to a subset of \( \{1/m, 2/m, \ldots, (m-1)/m\} \). For any true change point \( \tau \) in \( D_0 \), there will be a change point in \( \hat{D}^{(l)} \) that approaches \( \tau \) at a rate of \( n^{-2\beta/(2\beta+p)} \) up to some logarithmic factors. Second, it can be seen from the proof of Theorem 3.4.1 that the two error terms \( O\{n^{-2\beta/(2\beta+p)} \log^8 n\} \) and \( O(n^{-1/2}) \) in (iii) correspond to the bias and standard deviation of the proposed value estimator, respectively. When \( 2\beta > p \), the bias term is negligible. A Wald-type confidence interval can be constructed to infer \( V(d) \). The assumption \( 2\beta > p \) allow the deep learning estimator to converge at a rate faster than \( n^{-1/4} \). Such a condition is commonly imposed in the literature for inferring the average treatment effect (see e.g., Chernozhukov et al. 2017; Farrell et al. 2021). When \( \beta < p \), i.e., the underlying conditional mean function \( Q \) is not smooth enough, the proposed estimator suffers from a large bias and might converge at a rate that is slower than the usual parametric rate. This concern can be addressed by employing the A-learning method (see e.g., Murphy 2003b; Schulte et al. 2014; Shi et al. 2018). The A-learning method is more robust and requires weaker conditions to achieve the parametric rate. Specifically, it only requires the difference \( Q(x, 1) - Q(x, 0) \) to belongs to \( \Phi(\beta, c) \). This is weaker than requiring both \( Q(x, 1) \) and \( Q(x, 0) \) to belongs to \( \Phi(\beta, c) \). Third, to ensure the consistency of the proposed value estimator, we require that the distribution of the random variable \( d(X) \) induced by the target policy does not have point-masses at the change point locations. This condition is mild. For nondynamic policies where \( d(X) = d_0 \) almost surely, it requires \( d_0 \notin J(D_0) \). We remark that the set \( J(D_0) \) has a zero Lebesgue measure on \([0, 1] \). For dynamic policies, it automatically holds when \( d(X) \) has a bounded density on \([0, 1] \).

### 3.4.2 Properties under Model 2

We next consider Model 2 where the function \( Q(x, a) \) is continuous in the treatment space.

**Theorem 3.4.2.** Assume \( Q(x, a) \) is Lipschitz continuous, i.e., \( |Q(x, a_1) - Q(x, a_2)| \leq L|a_1 - a_2| \) for all \( a_1, a_2 \in [0, 1], x \in \mathcal{X}, \) and some constant \( L > 0 \). Assume (A1) and (A2), and \( m \) is proportional to \( n \) and \( \gamma_n \) is proportional to \( \max\{n^{-3/5}, n^{-2\beta/(2\beta+p)} \log^8 n\} \). Then for any target policy \( d \),

\[
\widehat{V}(d) - V(d) = O_p(n^{-1/5}) + O_p\{n^{-2\beta/(6\beta+3p)} \log^3 n\}.
\]

When \( 4\beta > 3p \), the convergence rate is given by \( O_p(n^{-1/5}) \). We remark that the above upper bound is valid for any target policy \( d \). The convergence rate in Theorem 3.4.2 may not be tight. To the best of our knowledge, no formal lower bounds of the value estimator
have been established in the literature in the continuous treatment setting. In the literature on multi-scale change point detection, there are lower bounds on the estimated time series (see e.g., Boysen et al. 2009). However, they considered settings without baseline covariates and it remains unclear how the rate of convergence of the estimated piecewise function can be translated into that of the value. We leave this for future research.

Finally, we clarify our theoretical contributions compared with the deep learning theory established in Farrell et al. (2021). First, Farrell et al. (2021) considered a single DNN, whereas we established the uniform convergence rate of several DNN estimators, since our proposal requires to train multiple DNN models. Establishing the uniform rate of convergence poses some unique challenges in deriving the results of Theorems 3.4.1 and 3.4.2. We need to control the initial number of the intervals \( m \) to be proportional to \( n \) and the order of penalty term \( \gamma_n \), so that uniform convergence rate can be established across all intervals. To address this difficulty, we derive the tail inequality to bound the rate of convergence of the DNN estimator and use the Bonferroni's correction to establish the uniform rate of convergence.

### 3.4.3 Comparison with Kernel-Based Methods

To simplify the analysis, we assume the kernel function is symmetric, the nuisance function estimators \( \hat{Q} \) and \( \hat{p} \) are set to their oracle values \( Q \) and \( b \), and that \( 4\beta > 3p \). Suppose Model 1 holds. In Appendix B.3, we show that the convergence rate of kernel-based methods is given by \( O_p(n^{-1/3}) \) with optimal bandwidth selection. In contrast, the proposed estimator converges at a faster rate of \( O_p(n^{-1/2}) \). Suppose Model 2 holds. In Appendix B.3, we show that the convergence rate of kernel-based methods is given by \( O_p(h) + O_p(n^{-1/2} h^{-1/2}) \). Thus, kernel-based estimators converge at a slower rate when the bandwidth undersmoothes or oversmoothes the data. In addition, as we have commented in Section 3.3.1, in cases where the second-order derivative of \( Q \) has an abrupt change in the treatment space, kernel-based methods suffer from either a large bias, or a large variance. Specifically, when \( h \) is either much larger than \( n^{-1/5} \) or much smaller than \( n^{-3/5} \), our estimator converges at a faster rate. Kernel-based estimators could converge at a faster rate when \( Q \) has a uniform degree of smoothness over the entire treatment space and the optimal bandwidth parameter is correctly identified.
3.5 Simulation Studies

In this section, we investigate the finite sample performance of the proposed method on the simulated and real datasets, in comparison to three kernel-based methods. The computing infrastructure used is a virtual machine in the AWS Platform with 72 processor cores and 144GB memory.

3.5.1 Simulation Settings

Simulated data are generated from the following model:
\[ Y | X, A \sim N \{ Q(X, A), 1 \}, \quad p(a|x) \sim \text{Uniform}[0, 1] \]  
where \( X = [X^{(1)}, \ldots, X^{(p)}] \). Consider the following different scenarios:

**S1**: \( Q(x, a) = (1 + x^{(1)})I(a < 0.35) + (x^{(1)} - x^{(2)})I(0.35 \leq a < 0.65) + (1 - x^{(2)})I(a \geq 0.65) \);

**S2**: \( Q(x, a) = I(a < 0.25) + \sin(2dx^{(1)})I(0.25 \leq a < 0.5) + \{0.5 - 8(x^{(1)} - 0.75)^2\}I(0.5 \leq a < 0.75) + 0.5I(a \geq 0.75) \);

**S3** *(toy example)*: \( Q(x, a) = 10\max\{a^2 - 0.25, 0\} \log(x^{(1)} + 2) \);

**S4**: \( Q(x, a) = 0.2(8 + 4x^{(1)} - 2x^{(2)} - 2x^{(3)}) - 2(1 + 0.5x^{(1)} + 0.5x^{(2)} - 2a)^2 \).

The function \( Q(x, a) \) is a piecewise function of \( a \) under Scenarios 1 and 2, and is continuous under Scenarios 3 (toy example considered in Section 3.3.1) and 4. We set the target policy to be the optimal policy that achieves the highest possible mean outcome. The dimension of the features is fixed to \( p = 20 \). We consider four choices of the sample size, corresponding to \( n = 50, 100, 200 \) or 300.

We compare the proposed DJL method with three kernel-based methods (Kallus and Zhou 2018; Colangelo and Lee 2020; Su et al. 2020b). In our implementation, we set \( \mathcal{A}_x \) to
the class of multilayer perceptrons (MLP) for each $\mathcal{S}$. This is a commonly used architecture in deep learning (Farrell et al. 2021). The optimization in (3.7) is solved via the MLP regressor implemented by Pedregosa et al. (2011) using a stochastic gradient descent algorithm, with tuning parameters set to the default values. In addition, we estimate the propensity score function using MLP as well. We set $m = n/10$ to achieve a good balance between the absolute error and the computational cost (see Figure B.1 in Appendix B.2 for details). The averaged computational time are summarized in Table B.1 with additional results under large sample sizes $n = 1000 \sim 10000$ in Table B.2, in Appendix B.2. Overall, it takes a few minutes (less than 1 min for $n = 50$ and 14 mins for $n = 300$) to implement DJL, whereas the runtime of Kallus and Zhou (2018)’s method is 365 mins for sample size $n = 50$ and over 48 hours for $n = 300$. Thus, as suggested in Kallus and Zhou (2018), to implement their method, we first compute $h^\ast$ using data with sample size $n_0 = 50$. To accommodate data with different $n$, we adjust $h^\ast$ by setting $h^\ast\{n_0/n\}^{0.2}$. To implement Colangelo and Lee (2020)’s estimator, we consider a list of bandwidths suggested in their paper, given by $h = c\sigma_A n^{-0.2}$ with $c \in \{0.5, 0.75, 1.0, 1.5\}$ and $\sigma_A$ is the standard deviation of the treatment. We then manually select the best bandwidth such that the resulting value estimator achieves the smallest mean squared error. The kernel-based method (SLOPE) by Su et al. (2020b) adopted the Lepski’s method for bandwidth selection. In their implementation, they used the IPW estimator to evaluate the value. For a fair comparison, we replace it with DR to make the resulting estimator more efficient.

The average estimated value and its standard deviation over 100 replicates are illustrated in Figure 3.2 for different methods, with detailed values reported in Table 2.1 in Appendix B.2. In addition, we provide the size of the final estimated partition under DJL in Table B.4 in Appendix B.2, which is much smaller than $m$ in most cases. It can be seen from Figure 3.2 that the proposed DJL method is very efficient and outperforms all competing methods in almost all cases. We note that the proposed method performs reasonably well even when the sample size is small ($n = 50$). In contrast, kernel-based methods fail to accurately estimate the value even in some cases when $n = 300$. Among the three kernel-based OPE approaches, we observe that the method developed by Su et al. (2020b) performs better in general. A potential limitation of our method is that it takes a longer computational time than the method of Colangelo and Lee (2020). To speed up the dynamic programming algorithm, for instance, the total variation or group-fused-lasso-type penalty can be used as a surrogate of the $L_0$ penalty to reduce the computational complexity (see e.g., Harchaoui and Lévy-Leduc 2010).
Table 3.2: The absolute error, the standard deviation, and the mean squared error of the estimated values under the optimal policy via different methods for the Warfarin data. The target value is given by $-0.278$.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Absolute error</th>
<th>Standard deviation</th>
<th>Mean squared error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Jump Learning</td>
<td>0.259</td>
<td>0.416</td>
<td>0.240</td>
</tr>
<tr>
<td>SLOPE</td>
<td>0.611</td>
<td>0.755</td>
<td>0.943</td>
</tr>
<tr>
<td>Kallus and Zhou (2018)</td>
<td>0.662</td>
<td>0.742</td>
<td>0.989</td>
</tr>
<tr>
<td>Colangelo and Lee (2020)</td>
<td>0.442</td>
<td>1.164</td>
<td>1.550</td>
</tr>
</tbody>
</table>

3.5.2 Real Data: Personalized Dose Finding

Warfarin is commonly used for preventing thrombosis and thromboembolism. We use the dataset provided by the International Warfarin Pharmacogenetics (Consortium 2009) for analysis. We choose $p = 81$ features considered in Kallus and Zhou (2018). This yields a total of 3964 with complete records. The outcome is defined as the absolute distance between the international normalized ratio (INR, a measurement of the time it takes for the blood to clot) after the treatment and the ideal value 2.5, i.e, $Y = -|\text{INR} - 2.5|$. We use the min-max normalization to convert the range of the dose level $A$ into $[0, 1]$. To compare among different methods, we calibrate the dataset to generate simulated outcomes. This allows us to use simulated data to calculate the bias and variance of each value estimator. Specifically, we first estimate the function $Q(x, a)$ via the MLP regressor using the whole dataset. The goodness-of-the-fit of the fitted model under the MLP regressor is reported in Table B.5 in Appendix B.2. We next use the fitted function $\hat{Q}(X, A)$ to simulate the data. For a given sample size $N$, we first randomly sample $N$ feature-treatment pairs $\{(a_j, x_j) : 1 \leq j \leq N\}$ from $\{(A_1, X_1), \cdots, (A_n, X_n)\}$ with replacement. Next, for each $j$, we generate the outcome $y_j$ according to $N\{\hat{Q}(x_j, a_j), \hat{\sigma}^2\}$, where $\hat{\sigma}$ is the standard deviation of the fitted residual $\{Y_i - \hat{Q}(X_i, A_i)\}_i$. This yields a simulated dataset $\{(x_j, a_j, y_j) : 1 \leq j \leq N\}$. We are interested in evaluating the mean outcome under the optimal policy as $d^*(X) \equiv \arg\max_{a \in [0,1]} Q(X, a)$.

We apply the proposed DJL method and the three kernel-based methods to the calibrated Warfarin dataset. Absolute errors, standard deviations, and mean squared errors of the estimated values under the optimal policy are reported in Table 3.2 over 20 replicates with different evaluation methods. It can be observed from Table 3.2 that our proposed DJL method achieves much smaller absolute error (0.259) and standard deviation (0.416) than the three kernel-based methods. The mean square error of the three competing estimators
are at least 3 times larger than DJL. The absolute error and standard deviation of Kallus and Zhou (2018)’s estimator and of the SLOPE in Su et al. (2020b) are approximately the same, due to that the bandwidth parameter is optimized. The estimator developed by Colangelo and Lee (2020)’s performs the worst. It suffers from a large variance, due to the suboptimal choice of the bandwidth. All these observations are aligned with the findings in our simulation studies.

3.6 Discussions

We proposed a brand-new OPE algorithm in continuous treatment domains. Combining our theoretical analysis and experiments, we are more confident that our proposed DJL method offers a practically much more useful policy evaluation tool compared to existing kernel-based approaches. There are some potential alternative directions to address the limitation of kernel-based approaches. Majzoubi et al. (2020) proposed a tree-based discretization to handle continuous actions in policy optimization for contextual bandits. Extending the tree-based discretization with adaptive pruning in OPE is a possible direction to handle our problem. Second, our proposed method can be understood as a special local kernel method with the boxcar kernel function, as we adaptively discretize the action space into a set of non-overlapping intervals. It would be practically interesting to investigate how to couple our procedure with general kernel functions.
CHAPTER

4

CALIBRATED OPTIMAL DECISION MAKING WITH MULTIPLE DATA SOURCES AND LIMITED OUTCOME

4.1 Introduction

In this chapter, we propose a new framework to handle heterogeneous samples and address the limited outcome simultaneously via a novel calibrated optimal decision making method, namely CODA. Motivated by the common data structures and similar data dependences in the MIMIC-III and eICU data, CODA naturally utilizes the common intermediate outcomes in multiple data sources via a calibration technique (Chen and Chen 2000; Chen 2002; Cao et al. 2009; Lumley et al. 2011), to improve the treatment decision rule by borrowing information from auxiliary samples.

Our contributions can be summarized as follows. First, to safely borrow information from auxiliary samples, we propose the comparable intermediate outcomes assumption, that is, the conditional means of intermediate outcomes given baseline covariates and
the treatment information are the same in the two samples. This assumption avoids the specification of the missing mechanism in auxiliary samples and is more practically reliable for heterogeneous studies. Second, all the current calibration-based methods require that covariates in the primary and auxiliary samples are from the same distribution. In this work, we allow baseline covariates across different studies to have either homogeneous or heterogeneous distributions. We propose a unified framework for deriving a calibrated doubly robust estimator of the conditional mean outcome of interest (known as the value function) through the projection onto the difference of value estimators for common intermediate outcomes in the two samples. When the distributions of baseline covariates differ in different samples, we construct the calibrated value estimator by rebalancing the value estimators of common intermediate outcomes in two samples based on their posterior sampling probability. Third, to handle the large-scale datasets (such as the MIMIC-III and eICU data) and obtain interpretable decision rules, we develop an iterative policy tree search algorithm to find the decision rule that maximizes the calibrated value estimator in the primary sample. The source code is publicly available at https://github.com/HengruiCai/CODA implemented in R language. Fourth, we establish the asymptotic normality of the calibrated value estimator under the estimated optimal IDR obtained by CODA for both homogeneous and heterogeneous covariates, which is shown to be more efficient than that obtained using the primary sample solely.

4.2 Statistical Framework

4.2.1 Notation and Formulation

For simplicity of exposition, we consider a study with two data sources. Suppose there is a primary sample of interest \(P\). Let \(X_P = [X_P^{(1)}, \ldots, X_P^{(r)}]^\top\) denote \(r\)-dimensional individual’s baseline covariates with the support \(X_P \in \mathbb{R}^r\), and \(A_P \in \{0, 1\}\) denote the binary treatment an individual receives. After a treatment \(A_P\) is assigned, we first obtain \(s\)-dimensional intermediate outcomes \(M_P = [M_P^{(1)}, \ldots, M_P^{(s)}]^\top\) with support \(M_P \in \mathbb{R}^s\), and then observe the primary outcome of interest \(Y_P\) with support \(Y_P \in \mathbb{R}\), the larger the better by convention. Denote \(N_P\) as the sample size for the primary sample, which consists of \(\{P_i = (X_{pi}, A_{pi}, M_{pi}, Y_{pi}), i = 1, \ldots, N_P\}\) independent and identically distributed across \(i\). To gain efficiency, we include an auxiliary sample \(U\) available from another source. The auxiliary sample \(U\) contains the same set of baseline covariates \(X_U = [X_U^{(1)}, \ldots, X_U^{(r)}]^\top\) (with the same ordering as \(X_P\) when \(r > 1\)), the treatment \(A_U\), and intermediate outcomes
with the support $X_U \in \mathbb{R}^r$, $\{0,1\}$, and $M_U \in \mathbb{R}^s$, respectively. Yet, the outcome of interest is limited in the primary sample and is not available in the auxiliary sample. Let $N_U$ denote the sample size for the independent and identically distributed auxiliary sample that includes $\{U_i = (X_{U,i}, A_{U,i}, M_{U,i}), i = 1, \ldots, N_U\}$. Denote $t = N_P/N_U$ as the sample ratio between the primary sample and the auxiliary sample, and $0 < t < +\infty$. And we allow the distributions of baseline covariates, treatments, and intermediate outcomes differ in two samples.

In the primary sample, define the potential outcomes $Y_p^*(0)$ and $Y_p^*(1)$ as the outcome of interest that would be observed after an individual receiving treatment 0 or 1, respectively. Similarly, we define the potential outcomes $\{M_p^*(0), M_p^*(1)\}$ and $\{M_U^*(0), M_U^*(1)\}$ as intermediate outcomes that would be observed after an individual receiving treatment 0 and 1 for the primary sample and the auxiliary sample, respectively. Define the propensity score function as the conditional probability of receiving treatment 1 given baseline covariates as $x$, denoted as $\pi_p(x) = \Pr(A_P = 1|X_P = x)$ for the primary sample and $\pi_U(x) = \Pr(A_U = 1|X_U = x)$ for the auxiliary sample. A decision rule is a deterministic function $d(\cdot)$ that maps covariate space $X_P$ to the treatment space $\{0,1\}$. Define the potential outcome of interest under $d(\cdot)$ as $Y_p^*(d) = Y_p^*(0)(1 - d(X_P)) + Y_p^*(1)d(X_P)$, which would be observed if a randomly chosen individual from the primary sample had received a treatment according to $d(\cdot)$, where we suppress the dependence of $Y_p^*(d)$ on $X_P$. The value function under $d(\cdot)$ is defined as the expectation of the potential outcome of interest over the primary sample as

$$V(d) = \mathbb{E}\{Y_p^*(d)\} = \mathbb{E}[Y_p^*(0)(1 - d(X_P)) + Y_p^*(1)d(X_P)].$$

As a result, the optimal decision rule (ODR) for the primary outcome of interest is defined as the maximizer of the value function among a class of decision rules $\Pi$,

$$d^{opt} = \arg\max_{d \in \Pi} V(d).$$

Similarly, we define the potential intermediate outcomes under $d(\cdot)$ for two samples as $M_p^*(d) = M_p^*(0)(1 - d(X_P)) + M_p^*(1)d(X_P)$ and $M_U^*(d) = M_U^*(0)(1 - d(X_U)) + M_U^*(1)d(X_U)$. Here, $M_P^*(d)$ and $M_U^*(d)$ are $s \times 1$ vectors if $s > 1$. 

$M_U = [M_U^{(1)}, \ldots, M_U^{(s)}]^\top$ (with the same ordering as $M_P$ when $s > 1$) as in the primary sample,
4.2.2 Assumptions

To identify ODR for the primary outcome of interest from observed data, as standard in the causal inference literature (Rubin 1978), we make the following assumptions:

(A1). Stable Unit Treatment Value Assumption:

\[
M_P = A_P M_P^*(1) + (1 - A_P) M_P^*(0); \quad Y_P = A_P Y_P^*(1) + (1 - A_P) Y_P^*(0);
M_U = A_U M_U^*(1) + (1 - A_U) M_U^*(0).
\]

(A2). Ignorability:

\[
\{M_P^*(0), M_P^*(1), Y_P^*(0), Y_P^*(1)\} \perp \perp A_P \mid X_P; \quad \{M_U^*(0), M_U^*(1)\} \perp \perp A_U \mid X_U.
\]

(A3). Positivity: There exist constants \(c_1, c_2, c_3, c_4\) such that with probability 1, \(0 < c_1 \leq \pi_P(x) \leq c_2 < 1\) for all \(x \in X_P\), and \(0 < c_3 \leq \pi_U(x) \leq c_4 < 1\) for all \(x \in X_U\).

The above assumptions (A1) to (A3) are standard in personalized decision-making (see Zhang et al. 2012b; Wang et al. 2018; Nie et al. 2020), to guarantee that the value function of intermediate outcomes in two samples and the value of the outcome of interest in the primary sample are estimable from observed data. Next, we make an assumption on the conditional means of intermediate outcomes in two samples to connect different data sources as follows.

(A4). Comparable Intermediate Outcomes Assumption:

\[
E(M_P | X_P = x, A_P = a) = E(M_U | X_U = x, A_U = a), \quad \text{for all } x \in X_P \cup X_U \text{ and } a \in \{0, 1\}.
\]

The above assumption states that the conditional means of intermediate outcomes given baseline covariates \(x\) and the treatment information \(a\) are the same in the two samples, for all \(x\) and \(a\) in the union of the supports of the two samples. This assumption automatically holds when the data sources are from the same population, where \(\{X_P, A_P, M_P\}\) has the same probability distribution of \(\{X_U, A_U, M_U\}\), as commonly assumed in the literature (see Yang and Ding 2019; Athey et al. 2020; Kallus and Mao 2020, and more details in Section 4.2.3). It is also worthy to mention that (A4) is testable based on observed data. For example, one can test the equality of two conditional mean models based on some posited parametric regression models such as linear regression (Chow 1960) or non-linear regression (Mahmoudi et al. 2018). In addition, we consider the class of decision rules \(\Pi\) that satisfies the following condition.
\(\text{(A5)}\). \(\Pi\) has a finite Vapnik-Chervonenkis dimension and is countable.

Assumption (A5) is commonly used in statistical learning and empirical process theory (see Kitagawa and Tetenov 2018; Rai 2018). When the baseline covariates have only a finite number of support points, any \(\Pi\) satisfies (A5). When the support of baseline covariates is continuous, assumption (A5) requires that \(\Pi\) is smaller than all measurable sets and can be well approximated by countably many elements. Popular classes of decision rules that satisfy (A5) include finite-depth decision trees (Athey and Wager 2017) and parametric decision rules (Zhang et al. 2012b). In this chapter, to handle the large-scale datasets such as the motivated MIMIC-III and eICU data and obtain interpretable decision rules, we consider a class of decision trees with finite-depth \(L_n \leq \kappa \log_2(n)\) for some \(\kappa < 1/2\), denoted as \(\Pi_1\), and search ODR within the class \(\Pi_1\).

4.2.3 Related Works

There are several recent works in using multiple data sources for estimating the average treatment effect (Yang and Ding 2019; Athey et al. 2020; Kallus and Mao 2020) or deriving a robust ODR to account for heterogeneity in multiple data sources (Shi et al. 018b; Mo et al. 2020). However, the settings and goals of these studies are different from what we consider here. Specifically, in the works of Yang and Ding (2019), Athey et al. (2020), and Kallus and Mao (2020), it was assumed that the two samples are from the same population and were linked together through a missing data framework, such as under the missing-at-random assumption. This allows to either develop a calibrated estimator using the common baseline covariates in both samples (Yang and Ding 2019) or impute the missing primary outcome in the auxiliary data (Athey et al. 2020; Kallus and Mao 2020), so that a more efficient estimator can be constructed for the average treatment effect. Whereas, the multiple data sources considered in our study may come from heterogeneous studies as in the MIMIC III and eICU data, and hence their missing data framework cannot be directly applied in our problem. For example, simply extending the calibration method for the average treatment effect considered in Yang and Ding (2019) may lead to a biased result when the covariate distributions in two samples are heterogeneous; while the adaption of the methods of Athey et al. (2020) and Kallus and Mao (2020) requires an untestable assumption that the conditional means of the outcome of interest given the baseline covariates, treatment, and intermediate outcomes are the same across samples. On the other hand, the main goal of Shi et al. (018b) and Mo et al. (2020) is to develop a single ODR that can work for multiple data sources with heterogeneity in data distributions or outcome models, and
their methods do not allow missingness in outcomes. In contrast, we are interested in safely improving the efficiency of the value estimator under ODR for the limited outcome, by leveraging available auxiliary data sources.

4.3 Method

4.3.1 Calibrated Optimal Decision Making for Homogeneous Baseline Covariates

We first focus on the case where the distributions of baseline covariates in different samples are the same, $X_P \sim X_U$. Consider the doubly robust estimators of the value functions (Zhang et al. 2012b) for the outcome of interest in the primary sample and intermediate outcomes in the two samples. Specifically, the doubly robust estimator for the outcome of interest in the primary sample is given by

$$
\hat{V}_P(d) = \frac{1}{N_P} \sum_{i=1}^{N_P} \frac{I(A_P = d(X_{Pi})) [Y_{Pi} - \hat{\mu}_P(X_{Pi}, d(X_{Pi}))]}{A_P \hat{\pi}_P(X_{Pi}) + (1 - A_P) \{1 - \hat{\pi}_P(X_{Pi})\}} + \hat{\mu}_P(X_{Pi}, d(X_{Pi})),$$

where $\hat{\pi}_P$ is the estimator of the propensity score function, and $\hat{\mu}_P(x, a)$ is the estimated conditional mean for $\mu_P(x, a) \equiv E(Y_P | X_P = x, A_P = a)$, in the primary sample. Following arguments in Zhang et al. (2012b); Luedtke and Van Der Laan (2016); Kitagawa and Tetenov (2018); Rai (2018), we have the asymptotic normality for the value estimator as

$$
\sqrt{N_P} \left\{ \hat{V}_P(d) - V(d) \right\} \rightsquigarrow N \left\{ 0, \sigma_Y^2(d) \right\}, \tag{4.1}
$$

where $\sigma_Y^2(d)$ is the asymptotic variance given any $d(\cdot)$. Next, we introduce the calibrated value estimator. By assumption (A4) and $X_P \sim X_U$, we establish the following lemma.

**Lemma 4.3.1.** Under assumptions (A1) - (A4) and $X_P \sim X_U$, we have

$$
E[M_p^*(d)] = E[M_p^*(0)\{1 - d(X_P)\} + M_p^*(1)d(X_P)]
= E[M_U^*(0)\{1 - d(X_U)\} + M_U^*(1)d(X_U)] = E[M_U^*(d)].
$$

The detailed proof of Lemma 4.3.1 is provided in the supplementary article. Based on Lemma 4.3.1, the value functions for the intermediate potential outcomes under $d(\cdot)$ in the
where $W_p(d), W_U(d), W(d)$ are $s \times 1$ value vectors when $s > 1$. This motivates us to derive the calibrated value estimator by projecting the value estimator of the outcome of interest in the primary sample on the differences of the value estimators of intermediate outcomes in the two samples. Following assumption (A4), we define the conditional mean of intermediate outcomes as $\theta(x, a) \equiv E(M_p|X_p = x, A_p = a) = E(M_U|X_U = x, A_U = a)$, which is a $s \times 1$ vector. Then, we have the doubly robust value estimators for intermediate outcomes in two samples as

$$
\bar{W}_p(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[ \frac{A_{p,i} - d(X_{p,i})}{A_{p,i}} \hat{\pi}_p(X_{p,i}) + (1 - A_{p,i}) \hat{\pi}_p(X_{p,i}) \right] + \hat{\theta}(X_{p,i}, d(X_{p,i}))
$$

$$
\bar{W}_U(d) = \frac{1}{N_U} \sum_{i=1}^{N_U} \left[ \frac{A_{U,i} - d(X_{U,i})}{A_{U,i}} \hat{\pi}_U(X_{U,i}) + (1 - A_{U,i}) \hat{\pi}_U(X_{U,i}) \right] + \hat{\theta}(X_{U,i}, d(X_{U,i}))
$$

where $\bar{W}_p(d)$ and $\bar{W}_U(d)$ are $s \times 1$ vectors, $\hat{\pi}_U$ is the estimator of the propensity score in the auxiliary sample, and $\hat{\theta}(x, a)$ is the estimated conditional mean function for $\theta(x, a)$ based on two samples combined under assumption (A4). Similarly, we have

$$
\sqrt{N_p} \left\{ \bar{W}_p(d) - W(d) \right\} \sim N_s \left\{ 0_s, \Sigma_p(d) \right\}, \sqrt{N_U} \left\{ \bar{W}_U(d) - W(d) \right\} \sim N_s \left\{ 0_s, \Sigma_U(d) \right\},
$$

where $0_s$ is the $s$-dimensional zero vector, $\Sigma_p$ and $\Sigma_U$ are $s \times s$ matrices presenting the asymptotic covariance matrices for two samples, and $N_s(\cdot, \cdot)$ is the $s$-dimensional multivariate normal distribution. Note that by Lemma 4.3.1, both $\bar{W}_p(d)$ and $\bar{W}_U(d)$ converge to the same value function $W(d)$. The following lemma establishes the asymptotic distribution of the differences of the value estimators of intermediate outcomes in the two samples, under some technical conditions (A6) and (A7) detailed in Section 4.4.

**Lemma 4.3.2.** Assume conditions (A1)-(A6) and (A7, i, ii, and iii) hold. Under $X_p \sim X_U$, we have

$$
\sqrt{N_p} \left\{ \bar{W}_p(d) - \bar{W}_U(d) \right\} \sim N_s \left\{ 0_s, \Sigma_M(d) \right\},
$$

where $\Sigma_M(d) = \Sigma_p(d) + T \Sigma_U(d)$ is an $s \times s$ asymptotic covariance matrix, and $T \equiv \lim_{N_p \to +\infty} t \in (0, +\infty)$. 

64
The key gradient to prove Lemma 4.3.2 lies in the fact that the two samples \((P, U)\) are collected from two different independent sources. The proof of Lemma 4.3.2 is given in the supplementary article. Based on Lemma 4.3.2, the asymptotic covariance matrix \(\Sigma_M(d)\) is a weighted sum of the asymptotic covariance from each sample, where the weight is determined by the limiting sample ratio between two samples. Based on the results established in (4.1) and Lemma 4.3.2, we have the following asymptotic joint distribution

\[
\sqrt{N_P} \left[ \frac{\hat{V}_P(d) - V(d)}{\hat{W}_P(d) - \hat{W}_U(d)} \right] \xrightarrow{N_{s+1}} \left[ 0_{s+1}, \begin{bmatrix} \sigma_Y^2(d), \rho(d) \end{bmatrix} \right], \text{ for all } d(\cdot),
\]

where \(\rho(d)\) is the \(s \times 1\) asymptotic covariance vector between the value estimator of the outcome of interest in the primary sample and the differences of the value estimators of intermediate outcomes between two samples. It follows that the conditional distribution of \(\sqrt{N_P} (\hat{V}_P(d) - V(d))\) given the estimated value differences of intermediate outcomes as

\[
\sqrt{N_P} \left[ \frac{\hat{V}_P(d) - V(d) - \rho(d)\Sigma_M^{-1}(d)(\hat{W}_P(d) - \hat{W}_U(d))}{\sqrt{N_P} (\hat{W}_P(d) - \hat{W}_U(d))} \right] \xrightarrow{N} \left( 0, \sigma_Y^2(d) - \rho(d)\Sigma_M^{-1}(d)\rho(d) \right), \text{ for all } d(\cdot). \tag{4.2}
\]

It can be observed from (4.2) that by projecting the value estimator of the outcome of interest on the estimated value differences of intermediate outcomes, we can achieve a smaller asymptotic variance. This result motivates us to construct the calibrated value estimator of \(V(d)\) as

\[
\hat{V}(d) = \hat{V}_P(d) - \hat{\rho}(d)\Sigma_M^{-1}(d)(\hat{W}_P(d) - \hat{W}_U(d)), \tag{4.3}
\]

where \(\hat{\rho}(d)\) is the estimator for \(\rho(d)\), and \(\hat{\Sigma}_M(d)\) is the estimator for \(\Sigma_M(d)\). These variances can be consistently estimated by a simple plug-in method without accounting for the variation in estimating nuisance functions, such as propensity scores and conditional mean models of outcomes, due to the rate double robustness properties. See the detailed estimation in Section C.1 of the supplementary article. Finally, the optimal decision rule under CODA for \(X_P \sim X_U\), namely CODA-HO, is to maximize the calibrated value estimator within a pre-specified class of decision rules \(\Pi\) as \(\hat{d} = \arg \max_{d \in \Pi} \hat{V}(d)\), with the corresponding estimated value function as \(\hat{V}(\hat{d})\).
4.3.2 Calibrated Optimal Decision Making for Heterogeneous Baseline Covariates

We next consider a more challenging case where the distributions of baseline covariates in the primary sample and the auxiliary sample are distinct, $X_p \not\sim X_U$. The results under Lemma 4.3.1 may not hold when the joint density of $X_p$ differs from the joint density of $X_U$. As such, we need to construct a new estimator of modified value differences such that it converges to a normal distribution with a zero mean even under $X_p \not\sim X_U$. To this end, we consider rebalancing the value estimators of common intermediate outcomes in two samples based on their posterior sampling probability. Specifically, we combine two samples together and denote the joint dataset as \( \{X_i, A_i, M_i, R_i, R_i Y_i\}_{i=1,\ldots,n} \), for $n = N_P + N_U$, where $R_i = 1$ if subject $i$ is from the primary sample and $R_i = 0$ if subject $i$ is from the auxiliary sample. Here, the distributions of baseline covariates, treatments, and intermediate outcomes are allowed to be different across different sub-samples, which distinguishes our work from the homogenous setting considered in the current literature (see Yang and Ding 2019; Athey et al. 2020; Kallus and Mao 2020).

To address the heterogeneous baseline covariates in two samples, also known as the covariate shift problem, a feasible way is to estimate the density functions of baseline covariates in two samples and adjust the corresponding estimator by the importance weights (see Sugiyama et al. 2007; Kallus 2021). However, we note that these methods cannot handle a relatively large number of baseline covariates due to the estimation of density functions, and can be hard to develop a simple inference procedure. Instead, we use a similar projection approach as developed in Section 4.3.1 to construct a new calibrated estimator through rebalancing to handle the heterogeneous baseline covariates and gain efficiency. To be specific, define the joint density of \( \{X_i, A_i, M_i\} \) given $R_i$ as $f(X_i = x, A_i = a, M_i = m | R_i = 1) \equiv f_P(x, a, m)$ and $f(X_i = x, A_i = a, M_i = m | R_i = 0) \equiv f_U(x, a, m)$, respectively, where $f_P(x, a, m)$ and $f_U(x, a, m)$ are the joint density function of \( \{X_P, A_P, M_P\} \) in the primary sample and the joint density function of \( \{X_U, A_U, M_U\} \) in the auxiliary sample, respectively.

By Bayesian theorem, we have the posterior sampling probability as

$$
\Pr(R_i = 1 | X_i = x, A_i = a, M_i = m) = \frac{\Pr(R_i = 1) f_P(x, a, m)}{\Pr(R_i = 1) f_P(x, a, m) + \Pr(R_i = 0) f_U(x, a, m)}.
$$

(4.4)

Here, we have $\Pr(R_i = 1) = \lim_{N_P \to \infty} N_P / (N_P + N_U) = \lim_{N_U \to \infty} t / (1 + t) = T / (1 + T)$. Based on (4.4), we can rebalance the value estimators of common intermediate outcomes in each sample to construct a new mean zero estimator. To this end, we estimate the posterior
where $\Sigma$ is built upon a more practical assumption. The proof of Lemma 4.3.3 can be found in the

not testable due to the unobserved outcome in the auxiliary sample. Therefore, our method

given

et al. 2020) require the missing at random assumption such that $R$ is independent of $Y$
given $(X, A, M)$. This implies $E(Y|X, A, M, R = 1) = E(Y|X, A, M, R = 0)$. This assumption is
not testable due to the unobserved outcome in the auxiliary sample. Therefore, our method
is built upon a more practical assumption. The proof of Lemma 4.3.3 can be found in
the supplementary article. It is immediate from Lemma 4.3.3 that $\hat{W}_i(d) - \hat{W}_0(d)$ is a mean zero
estimator. The following lemma establishes the asymptotic normality of the new estimator.

Lemma 4.3.3. Assume conditions (A1)-(A6) and (A7, i, iv, and v) hold. We have

$$\sqrt{n} \{ \hat{W}_i(d) - W^*(d) \} \rightsquigarrow N_s \{ 0, \Sigma_1(d) \}, \text{ and } \sqrt{n} \{ \hat{W}_0(d) - W^*(d) \} \rightsquigarrow N_s \{ 0, \Sigma_0(d) \},$$

where $\Sigma_1$ and $\Sigma_0$ are $s \times s$ asymptotic covariance matrices for each sub-sample, and

$$W^*(d) = \int E \{ M|d(X), X \} \{ Pr(R = 1)f(E, X) + Pr(R = 0)f(U, X) \} dX,$$

where $f(E, X)$ is the marginal density of baseline covariates in the primary sample and $f(U, X)$ is the marginal density of baseline covariates in the auxiliary sample.

Here, to show Lemma 4.3.3, we only require $E(M|X, A, R = 1) = E(M|X, A, R = 0)$, as indicated by assumption (A4). This is checkable since $(X, A, M)$ are observed in both samples. In contrast, current methods handling multiple datasets (see Kallus and Mao 2020; Athey et al. 2020) require the missing at random assumption such that $R$ is independent of $Y$
given $(X, A, M)$. This implies $E(Y|X, A, M, R = 1) = E(Y|X, A, M, R = 0)$. This assumption is
not testable due to the unobserved outcome in the auxiliary sample. Therefore, our method
is built upon a more practical assumption. The proof of Lemma 4.3.3 can be found in

67
Lemma 4.3.4. Suppose the conditions in Lemma 4.3.3 hold. We have
\[ \sqrt{n} \{ \widetilde{W}_1(d) - \widetilde{W}_0(d) \} \rightsquigarrow N_s \left\{ 0, \Sigma_R(d) \right\}, \]
where \( \Sigma_R(d) \) is a \( s \times s \) asymptotic covariance matrix.

Based on the results established in (4.1) and Lemma 4.3.4, we have the following asymptotic joint distribution
\[ \sqrt{N_p} \left[ \frac{\tilde{V}_p(d) - V(d)}{\sqrt{\frac{n}{N_p}} (\tilde{W}_1(d) - \tilde{W}_0(d))} \right] \rightsquigarrow N_{s+1} \left\{ 0_{s+1}, \begin{bmatrix} \sigma_Y^2(d), & \rho_R(d) \rho_R(d)^\top \end{bmatrix}, \begin{bmatrix} \rho_R(d), & \Sigma_R(d) \end{bmatrix} \right\}, \]
for all \( d(\cdot) \),
where \( \rho_R(d) \) is the \( s \times 1 \) asymptotic covariance vector between the value estimator of the outcome of interest in the primary sample and the new rebalanced value difference estimator of intermediate outcomes between two samples. It follows that the conditional distribution of \( \sqrt{N_p} \{ \tilde{V}_p(d) - V(d) \} \) given the estimated rebalanced value differences of intermediate outcomes as
\[ \sqrt{N_p} \left[ \tilde{V}_p(d) - V(d) - \sqrt{n/N_p} \rho_R(d)^\top \Sigma_R^{-1}(d) \{ \tilde{W}_1(d) - \tilde{W}_0(d) \} \right] \]
\[ \rightsquigarrow N \left\{ 0, \sigma_Y^2(d) - \rho_R(d)^\top \Sigma_R^{-1}(d) \rho_R(d) \right\}, \]
for all \( d(\cdot) \).

This yields the calibrated value estimator of \( V(d) \) under heterogeneous baseline covariates
\[ \tilde{V}_R(d) = \tilde{V}_p(d) - \sqrt{n/N_p} \tilde{\rho}_R(d)^\top \tilde{\Sigma}_R^{-1}(d) \{ \tilde{W}_1(d) - \tilde{W}_0(d) \}, \] (4.5)
where \( \tilde{\rho}_R(d) \) is the estimator for \( \rho_R(d) \), and \( \tilde{\Sigma}_R(d) \) is the estimator for \( \Sigma_R(d) \). Estimation on variances can be easily obtained using the simple plug-in method due to the rate double robustness as provided in Section C.1 of the supplementary article. Note that under the homogenous case where \( f_p(x, a, m) = f_U(x, a, m) \), according to (4.4), we have \( \Pr(R_i = 1 | X_i = x, A_i = a, M_i = m) = \Pr(R_i = 1) \). Then the above projection estimator will reduce to the estimator considered in Section 4.3.1. Therefore, the optimal decision rule under CODA for \( X_p \neq X_U \), namely CODA-HE, is to maximize the new calibrated value estimator \( \tilde{V}_R(d) \) within a pre-specified class of decision rules \( \Pi \) as \( \tilde{d}_R = \arg \max_{d \in \Pi} \tilde{V}_R(d) \), with the corresponding estimated value function as \( \tilde{V}_R(\tilde{d}_R) \).
4.3.3 Iterative Policy Tree Search Algorithm

We introduce the iterative policy tree search algorithm to implement CODA. We first elaborate on how to find the ODR that maximizes the calibrated value estimator for the homogenous case. Following the tree-based policy learning algorithm proposed in Athey and Wager (2017), we define the reward of the $i$-th individual in the primary sample as

$$\bar{v}_p^{(i)}(d) := \frac{\mathbb{I}[A_{P_i} = d(X_{P_i})][Y_{P_i} - \hat{\mu}_p \{X_{P_i}, d(X_{P_i})\}]}{A_{P_i}\hat{\pi}_p(X_{P_i}) + (1 - A_{P_i})\{1 - \pi_p(X_{P_i})\}} + \mu_p \{X_{P_i}, d(X_{P_i})\},$$

for $i \in \{1, \cdots, N_p\}$. Specifically, the reward of the $i$-th individual is $\bar{v}_p^{(i)}(1)$ under treatment 1 and $\bar{v}_p^{(i)}(0)$ under treatment 0. The decision tree allocates individuals to different treatments, and receives the corresponding rewards. The ODR based on the primary sample solely is obtained by maximizing the sum of these rewards through the exhaustive search within $\Pi_1$, denoted as $\tilde{d}_p = \arg \max_{d \in \Pi_1} \sum_{i=1}^{N_p} \bar{v}_p^{(i)}(d)$.

To develop ODR by CODA-HO within the class $\Pi_1$, based on (4.3), we construct the calibrated reward of the $i$-th individual in the primary sample by

$$\bar{w}_p^{(i)}(d) = \bar{v}_p^{(i)}(d) - \hat{\rho}(d)^\top \tilde{\Sigma}^{-1}_M(d)\{\mathbb{E}_p^{(i)}(d) - \bar{W}_d(d)\}, \quad (4.6)$$

where we define the value for the $i$-th individual in terms of intermediate outcomes as

$$\mathbb{E}_p^{(i)}(d) := \frac{\mathbb{I}[A_{P_i} = d(X_{P_i})][M_{P_i} - \hat{\theta} \{X_{P_i}, d(X_{P_i})\}]}{A_{P_i}\hat{\pi}_p(X_{P_i}) + (1 - A_{P_i})\{1 - \pi_p(X_{P_i})\}} + \theta \{X_{P_i}, d(X_{P_i})\}.$$

Here, notice that the sample mean of (4.6) over the primary sample yields (4.3). Therefore, the decision tree that maximizes the sum of rewards defined in (4.6) also maximizes the calibrated value estimator in (4.3). The finite-depth decision tree-based ODR under CODA-HO is then defined by

$$\tilde{d} = \arg \max_{d \in \Pi_1} \sum_{i=1}^{N_p} \bar{v}_p^{(i)}(d). \quad (4.7)$$

Yet, the estimators $\hat{\rho}(d)$ and $\tilde{\Sigma}_M^{-1}(d)$ defined in (4.6) are calculated using two samples’ information (see details in Section C.1 of the supplementary article) given a specific decision rule $d$, and thus the tree-based policy learning in Athey and Wager (2017) is not directly applicable to solve (4.7). To address this difficulty, we propose an iterative policy tree search algorithm consisting of four steps as follows.

**Step 1.** Find the ODR based on the primary sample solely ($\tilde{d}_p$) as an initial decision tree.
Step 2. Estimate \( \rho(\cdot) \) and \( \Sigma_M(\cdot) \) by plugging in \( d = \hat{d}_p \). Thus, the calibrated reward for the \( i \)-th individual can be approximated by 
\[
\hat{v}_p^{(i)}(1) - \rho(\hat{d}_p)^	op \Sigma_M^{-1}(\hat{d}_p)\{\hat{w}_p^{(i)}(1) - \hat{W}_U(1)\}
\]
under treatment 1, and 
\[
\hat{v}_p^{(i)}(0) - \rho(\hat{d}_p)^	op \Sigma_M^{-1}(\hat{d}_p)\{\hat{w}_p^{(i)}(0) - \hat{W}_U(0)\}
\]
under treatment 0.

Step 3. Search for the optimal decision tree within the class \( \Pi_1 \) to achieve a maximum overall calibrated reward, denoted as \( \hat{d}^{(1)} \). This step can be solved by applying the tree-based policy learning in Athey and Wager (2017) with an updated reward matrix.

Step 4. Repeat steps 2 and 3 for \( k = 1, \cdots, K \), by replacing the previous estimated decision tree \( \hat{d}^{(k-1)} \) (\( \hat{d}^{(0)} = \hat{d}_p \)) with the new estimated decision tree \( \hat{d}^{(k)} \) until it’s convergent or achieves the maximum number of iterations \( K \). It is observed in our simulation studies (see Section 4.5) that \( \hat{d}_p \) is fairly close to \( \hat{d} \), and thus one iteration is usually sufficient to find ODR under CODA in practice.

The above iterative policy search algorithm can be extended to the heterogeneous case as well as parametric decision rules, as provided in Section C.2 of the supplementary article.

### 4.4 Theoretical Properties

In this section, we investigate the theoretical properties of the value estimator under CODA in (4.3) and (4.5), corresponding to two cases where the distributions of baseline covariates are the same or different, respectively. All the proofs are provided in the supplementary article. As standard in statistical inference for personalized decision-making (Zhang et al. 2012b; Luedtke and Van Der Laan 2016; Kitagawa and Tetenov 2018; Rai 2018), we introduce the following conditions to derive our theoretical results.

(A6) Suppose the supports \( X_p, X_U, M_p, M_U \), and \( Y_p \) are bounded.

(A7) Rate double robustness for model misspecification: for \( a = 0, 1, \)

(i) \( [E_{X \in X_p} \{\mu_p(X, a) - \hat{\mu}_p(X, a)\}^2 \{\pi_p(X) - \hat{\pi}_p(X)\}^2]^\frac{1}{2} = o_p(N_p^{-1/2}) \),

(ii) \( [E_{X \in X_p} \{\theta(X, a) - \hat{\theta}(X, a)\}^2 \{\pi_p(X) - \hat{\pi}_p(X)\}^2]^\frac{1}{2} = o_p(N_p^{-1/2}) \),

(iii) \( [E_{X \in X_U} \{\theta(X, a) - \hat{\theta}(X, a)\}^2 \{\pi_U(X) - \hat{\pi}_U(X)\}^2]^\frac{1}{2} = o_p(N_U^{-1/2}) \),

(iv) \( [E_{X \in X_p, \cup X_p} \{\theta(X, a) - \hat{\theta}(X, a)\}^2 \{\pi(X) - \hat{\pi}(X)\}^2]^\frac{1}{2} = o_p(n^{-1/2}) \),

(v) \( [E_{X \in X_p, \cup X_p, M \in M_p, \cup M_U} \{\theta(X, a) - \hat{\theta}(X, a)\}^2 \{r(X, a, M) - \bar{r}(X, a, M)\}^2]^\frac{1}{2} = o_p(n^{-1/2}) \).

(A8) There exist some constants \( \gamma, \lambda > 0 \) such that

\[
\Pr\{0 < |E(Y_p|X_p, A_p = 1) - E(Y_p|X_p, A_p = 0)| \leq \xi\} = O(\xi^\gamma),
\]
where the big-\(O\) term is uniform in \(0 < \xi \leq \lambda\).

Assumption (A6) is a technical assumption sufficient to establish the uniform convergence results. A similar assumption is frequently used in the literature of optimal treatment regime estimation (see Zhang et al. 2012b; Zhao et al. 2012b; Zhou et al. 2017). Assumption (A7) (i)-(iv) requires the estimated conditional mean outcomes and propensity score functions to converge at certain rates in each decision-making problem (for the primary outcome of interest \(Y_P\), the intermediate outcomes \(M_P\) and \(M_U\), and the joint intermediate outcomes). This assumption is commonly imposed in the causal inference literature (see Athey et al. 2020; Kallus and Mao 2020) to derive the asymptotic distribution of the estimated average treatment effect with either parametric or non-parametric estimators (see Wager and Athey 2018; Farrell et al. 2021). We extend it to (A7) (v) that the estimators of the posterior sampling probability and conditional mean functions of intermediate outcomes converge at certain rates, by which together with (A7) (iv) one can establish the asymptotic distribution of the rebalanced value estimators based on the joint sample. Finally, assumption (A8) is well known as the margin condition, which is often adopted in the literature to derive a sharp convergence rate for the value function under the estimated optimal decision rule (see Qian and Murphy 2011; Luedtke and Van Der Laan 2016; Kitagawa and Tetenov 2018). We first establish the consistency of the proposed estimators in the following theorem.

**Theorem 4.4.1.** (Consistency) Suppose conditions (A1)-(A7) hold. We have

(i) \(\Delta^2_Y(d) = \sigma^2_Y(d) + o_p(1)\); (ii) \(\Delta^0(d) = \rho(d) + o_p(1)\); (iii) \(\Sigma_M(d) = \Sigma_M(d) + o_p(1)\);

(iv) \(\Delta^0_R(d) = \rho_R(d) + o_p(1)\); (v) \(\Sigma_R(d) = \Sigma_R(d) + o_p(1)\);

(vi) if \(X_P \sim X_U\), \(\Delta(d) = V(d) + o_p(1)\); (vii) \(\Delta_R(d) = V(d) + o_p(1)\).

Note that the result (vi) in Theorem 4.4.1 requires additional homogeneous baseline covariates, while the rest results hold for either the homogeneous or the heterogeneous case. We remark that the key step of the proof is to decompose the variance estimators based on the true value estimator defined at the individual level in each sample. This allows replacing the estimators with their true models based on the rate doubly robustness in assumption (A7) with a small order. We next show the asymptotic normality of the proposed calibrated value estimator in the following theorem for homogeneous baseline covariates.

**Theorem 4.4.2.** (Asymptotic Normality for Homogeneous Baseline Covariates) Suppose \(\{d^{opt}, \hat{d}\} \in \Pi_1\) or \(\{d^{opt}, \hat{d}\} \in \Pi_2\). Assume conditions (A1)-(A6), (A7. i, ii, and iii), and (A8)
hold. Under $X_p \sim X_U$, we have

$$\sqrt{N_p} \left\{ \hat{V}(\hat{d}) - V(d^{opt}) \right\} \rightarrow N \left\{ 0, \sigma^2(d^{opt}) \right\},$$

where $\sigma^2(d^{opt}) = \sigma^2_Y(d^{opt}) - \rho(d^{opt})^\top \Sigma_d^{-1}(d^{opt}) \rho(d^{opt})$.

The condition in Theorem 4.4.2 that ${d^{opt}, \hat{d}} \in \Pi_1$ or ${d^{opt}, \hat{d}} \in \Pi_2$ requires the true ODR falls into the class of decision rules of interest, such that the resulting estimated decision rule by CODA-HO is not far away from the true ODR. The proof of Theorem 4.4.2 consists of three steps. We first replace the estimated propensity score function and the estimated conditional mean function in $\hat{V}(\hat{d})$ by their counterparts based on the rate doubly robustness in assumption (A7) with a small order $o_p(N_p^{-1/2})$. Secondly, we show the value estimator under the estimated decision rule by CODA-HO converges to the value estimator under the true ODR at a rate of $o_p(N_p^{-1/2})$, based on the empirical process theory and the margin condition in (A8). The proof is nontrivial to the literature of personalized decision-making, by noticing that the estimated decision rule by CODA-HO is to maximize the newly proposed calibrated value estimator. Lastly, the asymptotic normality follows the central limit theorem. Since the two samples ($P$ and $U$) are independently collected from two separate studies, we can explicitly derive the asymptotic variance of the calibrated value estimator under the estimated decision rule by CODA-HO. According to the results in Theorem 4.4.2, when $\rho(d^{opt})$ is a non-zero vector, we have $\sigma^2_Y(d^{opt}) - \rho(d^{opt})^\top \Sigma_d^{-1}(d^{opt}) \rho(d^{opt}) < \sigma^2_Y(d^{opt})$, since $\Sigma_d(d^{opt})$ is positive definite. In other words, if the primary outcome of interest is correlated with one of the selected intermediate outcomes, the asymptotic variance of the calibrated value estimator under CODA-HO is strictly smaller than the asymptotic variance of the value estimator under the ODR obtained based on the primary sample solely. The larger the correlation is, the smaller variance we can achieve. Hence, the proposed calibrated value estimator is more efficient by integrating different data sources.

Based on Theorem 4.4.2, By plugging the estimates $\hat{\sigma}_Y^2(\hat{d}), \hat{\rho}(\hat{d}), \Sigma_d(\hat{d})$, the asymptotic variance of $\hat{V}(\hat{d})$ can be consistently estimated by $\hat{\sigma}_Y^2(\hat{d}) := \hat{\sigma}_Y^2(\hat{d}) - \hat{\rho}(\hat{d})^\top \Sigma_d^{-1}(\hat{d}) \hat{\rho}(\hat{d})$. Therefore, a two-sided $1 - \alpha$ confidence interval for $V(d^{opt})$ under CODA-HO is given by

$$\left[ \hat{V}(\hat{d}) - z_{\alpha/2} \hat{\sigma} / \sqrt{N_p}, \hat{V}(\hat{d}) + z_{\alpha/2} \hat{\sigma} / \sqrt{N_p} \right],$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$-th quantile of a standard normal distribution. Similarly, we establish the asymptotic normality for the heterogeneous case as follows.

Theorem 4.4.3. (Asymptotic Normality for Heterogeneous Baseline Covariates) Suppose
\( \{d^{opt}, \hat{d}_R\} \in \Pi_1 \) or \( \{d^{opt}, \hat{d}_R\} \in \Pi_2 \). Under conditions (A1)-(A6), (A7. i, iv, and v), and (A8), we have
\[
\sqrt{N_P} \left\{ \widehat{V}_R(\hat{d}_R) - V(d^{opt}) \right\} \rightsquigarrow N \left\{ 0, \sigma^2_R(d^{opt}) \right\},
\]
where \( \sigma^2_R(d^{opt}) = \sigma^2_Y(d^{opt}) - \rho_R(d^{opt})^\top \Sigma_R^{-1}(d^{opt}) \rho_R(d^{opt}) \).

Here, we also require the true ODR falls into the class of decision rules of interest, such that the resulting estimated decision rule by CODA-HE is not far away from the true ODR. The proof of Theorem 4.4.3 is similar to Theorem 4.4.2. When \( \rho_R(d^{opt}) \) is a non-zero vector, we have
\[
\sigma^2_Y(d^{opt}) - \rho_R(d^{opt})^\top \Sigma_R^{-1}(d^{opt}) \rho_R(d^{opt}) < \sigma^2_Y(d^{opt}) \text{ such that the proposed calibrated value estimator is more efficient.}
\]
The corresponding two-sided \( 1 - \alpha \) confidence interval for \( V(d^{opt}) \) under CODA-HE is given by
\[
\left[ \widehat{V}_R(\hat{d}_R) - z_{\alpha/2} \hat{\sigma}_R / \sqrt{N_P} , \quad \widehat{V}_R(\hat{d}_R) + z_{\alpha/2} \hat{\sigma}_R / \sqrt{N_P} \right], \tag{4.9}
\]
where \( \hat{\sigma}_R = \sigma^2_Y(\hat{d}_R) - \rho_R(\hat{d}_R)^\top \Sigma_R^{-1}(\hat{d}_R) \rho_R(\hat{d}_R) \) is the estimator of \( \sigma^2_R(d^{opt}) \).

### 4.5 Simulation Studies

#### 4.5.1 Evaluation on Calibrated Value Estimator for Homogeneous Baseline Covariates

In the following, we produce the primary sample and the auxiliary sample. Suppose their baseline covariates \( X = [X^{(1)}, \cdots, X^{(r)}]^\top \), the treatment \( A \), and intermediate outcomes \( M = [M^{(1)}, \cdots, M^{(s)}]^\top \) are generated from the following model:
\[
A \sim \text{Bernoulli}[\pi(X)], \quad X^{(1)}, \cdots, X^{(r)} \sim \text{Uniform}[-2, 2],
\]
\[
M = U^M(X) + AC^M(X) + \epsilon^J, \quad J \in \{P, U\}
\]
where \( U^M(\cdot) \) is the baseline function of intermediate outcomes, \( C^M(\cdot) \) is the contrast function that describes the treatment-covariates interaction effects for intermediate outcomes, and \( \epsilon^M \) is the random error. Here, we consider a logistic regression for the propensity score, \( \logit[\pi(X)] = 0.4 + 0.2X^{(1)} - 0.2X^{(2)} \). The outcome of interest is generated in the primary sample only, according to the following regression model:
\[
Y = U^Y(X) + AC^Y(X) + \epsilon^Y,
\]
where $U^Y(\cdot)$ is the baseline function for the outcome of interest, $C^Y(\cdot)$ is the contrast function that describes the treatment-covariates interaction effects for the outcome of interest, and $\epsilon^Y$ is the random error. We consider the case where the intermediate outcomes in the two samples are generated from different distributions to account for their heterogeneity. Specifically, the noises of intermediate outcomes in the auxiliary sample are set to be $\epsilon^U \sim \text{Uniform}[-1, 1]$. In addition, we let the noises of intermediate outcomes in the primary sample $\epsilon^E$ and the noise of the outcome of interest in the primary sample $\epsilon^Y$ generated from a bivariate normal distribution with mean zero, variance vector as $[2, 1.5]$, and a positive correlation of 0.7. Note that given the baseline covariates and the treatment, the conditional means of intermediate outcomes in the two samples are the same, and thus assumption (A4) is satisfied. We set the dimension of covariates as $r = 2$ and the dimension of the intermediate outcome as $s = 1$, and consider the following two scenarios respectively.

**Scenario 1 (decision tree):**

\[
\begin{aligned}
U^M(X) &= X^{(1)} + 2X^{(2)}, \\
C^M(X) &= X^{(1)} \times X^{(2)}; \\
U^Y(X) &= 2X^{(1)} + X^{(2)}, \\
C^Y(X) &= 2X^{(1)} \times X^{(2)}. \\
\end{aligned}
\]

**Scenario 2 (linear rule):**

\[
\begin{aligned}
U^M(X) &= X^{(1)} + 2X^{(2)}, \\
C^M(X) &= X^{(1)} - X^{(2)}; \\
U^Y(X) &= 2X^{(1)} + X^{(2)}, \\
C^Y(X) &= 2\{X^{(2)} - X^{(1)}\}. \\
\end{aligned}
\]

For Scenario 1, the true ODR can be represented by a decision tree as $d^{opt}(X) = I\{X^{(1)}X^{(2)} > 0\}$, which is unique up to permutation. Its true value $V(d^{opt})$ can be calculated by Monte Carlo approximations, as 0.999. In Scenario 2, the true ODR takes a form of a linear rule as $d^{opt}(X) = I\{X^{(2)} - X^{(1)} > 0\}$, with its true value $V(d^{opt})$ as 1.333. Since $d^{opt}$ in Scenario 2 cannot be represented by a decision tree, $d^{opt} \notin \Pi_1$, we have $\max_{d\in\Pi_1} V(d) = 1.251$ for Scenario 2 as the true value under the optimal decision tree, which is smaller than $V(d^{opt})$.

We consider $N_U = 2000$ and allow $N_P$ chosen from the set $\{500, 1000\}$. For each setting, we conduct 500 replications. Here, we search the ODR using CODA-HO based on two samples and the ODR based on the primary sample solely, within the class of decision trees $\Pi_1$. To illustrate the approximation error due to policy search, we directly plug the true decision rule $d^{opt}$ in the value estimators for comparison. The empirical results are reported in Tables 4.1 and 4.2, for Scenarios 1 and 2, respectively, aggregated over 500 replications. We summarize the true value function $V(\cdot)$ of a given decision rule computed using the Monte Carlo simulation method, the estimated value $\hat{V}(\cdot)$ with its standard deviation $SD\{\hat{V}(\cdot)\}$,
Table 4.1: Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenario 1.

<table>
<thead>
<tr>
<th>Method (Rule)</th>
<th>CODA ($d^{opt}$)</th>
<th>CODA ($d$)</th>
<th>ODR ($d^{opt}$)</th>
<th>ODR ($d_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p =$</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
</tr>
<tr>
<td>True Value $V(\cdot)$</td>
<td>0.999 0.999</td>
<td>0.963 0.976</td>
<td>0.999 0.999</td>
<td>0.967 0.976</td>
</tr>
<tr>
<td>Estimated Value $\hat{V}(\cdot)$</td>
<td>0.998 0.996</td>
<td>1.053 1.030</td>
<td>1.006 0.994</td>
<td>1.095 1.050</td>
</tr>
<tr>
<td>$SD{\hat{V}(\cdot)}$</td>
<td>0.124 0.098</td>
<td>0.123 0.098</td>
<td>0.173 0.127</td>
<td>0.171 0.125</td>
</tr>
<tr>
<td>$E{\tilde{\sigma}}$</td>
<td>0.129 0.096</td>
<td>0.129 0.095</td>
<td>0.182 0.128</td>
<td>0.181 0.128</td>
</tr>
<tr>
<td>Coverage Probabilities</td>
<td>95.6% 95.0%</td>
<td>94.6% 94.6%</td>
<td>96.8% 95.2%</td>
<td>94.2% 94.4%</td>
</tr>
<tr>
<td>Improved Efficiency</td>
<td>29.1% 25.0%</td>
<td>28.7% 25.8%</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>$\hat{\rho}(\cdot)$</td>
<td>12.4 12.3</td>
<td>12.4 12.3</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>$\Sigma_M(\cdot)$</td>
<td>18.8 20.8</td>
<td>18.8 20.8</td>
<td>/ /</td>
<td>/ /</td>
</tr>
</tbody>
</table>

Table 4.2: Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenario 2, where $\max_{d \in \Pi_1} V(d) = 1.251$.

<table>
<thead>
<tr>
<th>Method (Rule)</th>
<th>CODA ($d^{opt}$)</th>
<th>CODA ($d$)</th>
<th>ODR ($d^{opt}$)</th>
<th>ODR ($d_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p =$</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
</tr>
<tr>
<td>True Value $V(\cdot)$</td>
<td>1.333 1.332</td>
<td>1.236 1.239</td>
<td>1.333 1.331</td>
<td>1.227 1.323</td>
</tr>
<tr>
<td>Estimated Value $\hat{V}(\cdot)$</td>
<td>1.327 1.321</td>
<td>1.321 1.303</td>
<td>1.329 1.311</td>
<td>1.350 1.319</td>
</tr>
<tr>
<td>$SD{\hat{V}(\cdot)}$</td>
<td>0.110 0.085</td>
<td>0.108 0.085</td>
<td>0.156 0.112</td>
<td>0.154 0.110</td>
</tr>
<tr>
<td>$E{\tilde{\sigma}}$</td>
<td>0.115 0.085</td>
<td>0.116 0.086</td>
<td>0.162 0.114</td>
<td>0.161 0.114</td>
</tr>
<tr>
<td>Coverage Probabilities</td>
<td>95.8% 95.2%</td>
<td>96.2% 94.6%</td>
<td>95.4% 95.2%</td>
<td>96.4% 95.2%</td>
</tr>
<tr>
<td>Improved Efficiency</td>
<td>29.0% 25.4%</td>
<td>28.0% 24.6%</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>$\hat{\rho}(\cdot)$</td>
<td>10.7 10.6</td>
<td>10.6 10.5</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>$\Sigma_M(\cdot)$</td>
<td>17.8 19.4</td>
<td>17.8 19.5</td>
<td>/ /</td>
<td>/ /</td>
</tr>
</tbody>
</table>

Based on Table 4.1 and 4.2, it is clear that the proposed CODA-HO method is more efficient than the ODR method based on the primary sample solely, in all cases. To be specific, CODA-HO improves efficiency by 28.7% in Scenario 1 and 28.0% in Scenario 2.

Based on Tables 4.1 and 4.2, it is clear that the proposed CODA-HO method is more efficient than the ODR method based on the primary sample solely, in all cases. To be specific, CODA-HO improves efficiency by 28.7% in Scenario 1 and 28.0% in Scenario 2.
for $N_p = 500$, and by 25.8% in Scenario 1 and 24.6% in Scenario 2 for $N_p = 1000$. On the other hand, the values under CODA-HO approach the true as the sample size $N_p$ increases in all scenarios. Specifically, the proposed method achieves $V(\hat{d}) = 0.977$ in Scenario 1 ($V(d^{opt}) = 0.999$) and $V(\hat{d}) = 1.240$ in Scenario 2 ($\max_{d \in \Pi} V(d) = 1.251$) when $N_p = 1000$. These results are comparable to or slightly better than the values under the estimated ODR based on the primary sample solely.

Two findings help to verify Theorem 4.4.2. First, the mean of the estimated standard error of the value function ($\mathbb{E}[\tilde{\sigma}]$) is close to the standard deviation of the estimated value ($SD\{\hat{V}\}$), and gets smaller as the sample size $N_p$ increases. Second, the empirical coverage probabilities of the proposed 95% confidence interval in (4.8) approach to the nominal level in all settings. All these findings are further justified by directly applying the true optimal decision rule into the proposed calibrated value estimator. It can be observed in Tables 4.1 and 4.2 that the estimated asymptotic correlation $\hat{\rho}$ and the estimated asymptotic covariance $\hat{\Sigma}_M$ under the estimated decision rule by CODA-HO are very close to that under the true rule $d^{opt}$, with only one iteration. This supports the implementation technique discussed in Section 4.3.3. Additional simulation to demonstrate the improved efficiency using CODA in various scenarios with high-dimensional covariates and multiple intermediate outcomes are provided in Section C.3 of the supplementary article.

### 4.5.2 Evaluation on Calibrated Value Estimator for Heterogeneous Baseline Covariates

We next consider samples with heterogeneous baseline covariates generated by:

\[
A_P \sim \text{Bernoulli}\{\pi(X_P)\}, \quad X_{P}^{(1)}, \cdots, X_{P}^{(r)} \sim \text{Uniform}\{-2, 2\}, \\
A_U \sim \text{Bernoulli}\{\pi(X_U)\}, \quad X_{U}^{(1)}, \cdots, X_{U}^{(r)} \sim \text{Uniform}\{-1, 1.5\}, \\
M_P = U^M(X_P) + A_P C^M(X_P) + \epsilon^E, \quad M_U = U^M(X_U) + A_U C^M(X_U) + \epsilon^U.
\]

Here, we consider the same logistic regression for the propensity score as $\text{logit}\{\pi(X)\} = 0.4 + 0.2X^{(1)} - 0.2X^{(2)}$. Similar to Section 4.5.1, we generate the outcome of interest for the primary sample only by $Y_P = U^Y(X_P) + A_P C^Y(X_P) + \epsilon^Y$. The noises of intermediate outcomes in the auxiliary sample are set to be $\epsilon^U \sim \text{Uniform}\{-1, 1\}$, while $\epsilon^E$ and $\epsilon^Y$ are generated from a bivariate normal distribution with mean zero, variance vector as $[2, 1.5]$, and a positive correlation of 0.7.

We set the dimension of covariates as $r = 2$ and the dimension of the intermediate
Table 4.3: Empirical results of the proposed CODA-HE method in comparison to the ODR based on the primary sample solely under Scenario 1 with heterogeneous baseline covariates.

<table>
<thead>
<tr>
<th>Method (Rule)</th>
<th>CODA ($d^{\text{o}p}t$)</th>
<th>CODA ($d_{R}$)</th>
<th>ODR ($d^{\text{o}p}t$)</th>
<th>ODR ($d_{p}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p =$</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
</tr>
<tr>
<td>True Value $V(\cdot)$</td>
<td>0.999 0.975 0.984</td>
<td>0.999 0.967 0.977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated Value $\hat{V}(\cdot)$</td>
<td>1.001 0.992 1.059 1.029</td>
<td>1.006 0.993 1.095 1.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SD{\hat{V}(\cdot)}$</td>
<td>0.162 0.112 0.161 0.112</td>
<td>0.173 0.122 0.171 0.121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E{\hat{\sigma}}$</td>
<td>0.172 0.120 0.171 0.120</td>
<td>0.182 0.128 0.181 0.128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage Probabilities</td>
<td>96.8% 95.8% 97.0% 94.8%</td>
<td>96.8% 96.0% 94.2% 94.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved Efficiency</td>
<td>5.5% 6.3% 5.5% 6.3%</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho(\cdot)$</td>
<td>2.43 3.08 2.39 3.06</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{M}(\cdot)$</td>
<td>3.19 4.67 3.17 4.65</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: Empirical results of the proposed CODA-HE method in comparison to the ODR based on the primary sample solely under Scenario 2 with heterogeneous baseline covariates, where $\max_{d \in \Pi_1} V(d) = 1.251$.

<table>
<thead>
<tr>
<th>Method (Rule)</th>
<th>CODA ($d^{\text{o}p}t$)</th>
<th>CODA ($d_{R}$)</th>
<th>ODR ($d^{\text{o}p}t$)</th>
<th>ODR ($d_{p}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p =$</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
<td>500 1000</td>
</tr>
<tr>
<td>True Value $V(\cdot)$</td>
<td>1.333 1.232 1.239</td>
<td>1.333 1.226 1.235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated Value $\hat{V}(\cdot)$</td>
<td>1.325 1.320 1.317 1.288</td>
<td>1.329 1.322 1.350 1.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SD{\hat{V}(\cdot)}$</td>
<td>0.140 0.103 0.139 0.102</td>
<td>0.156 0.118 0.154 0.116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E{\hat{\sigma}}$</td>
<td>0.148 0.104 0.148 0.104</td>
<td>0.162 0.114 0.161 0.114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage Probabilities</td>
<td>96.0% 95.0% 95.8% 95.6%</td>
<td>95.4% 93.6% 94.2% 93.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved Efficiency</td>
<td>8.6% 8.8% 8.1% 8.8%</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho(\cdot)$</td>
<td>2.72 3.44 2.68 3.39</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{M}(\cdot)$</td>
<td>3.56 5.18 3.52 5.12</td>
<td>/ / / /</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

outcome as $s = 1$, and consider Scenarios 1 and 2 in Section 4.5.1. Using a similar procedure introduced in Section 4.5.1, we apply the proposed CODA-HE method ($\hat{d}_{R}$), in comparison to the ODR method based on the primary sample solely ($\hat{d}_{p}$). The empirical results are summarized in Tables 4.3 and 4.4, for Scenarios 1 and 2, respectively, aggregated over 500 replications. The coverage probabilities are calculated based on the 95% confidence interval in (4.9).

It can be observed from Tables 4 and 5 that the proposed CODA-HE method is more efficient than the ODR method based on the primary sample solely, in all scenarios, under heterogeneous baseline covariates. Specifically, CODA-HE improves efficiency by 6.3% in Scenario 1 and 8.8% in Scenario 2 for $N_p = 1000$. In addition, the values under CODA-HE
approach the true as the sample size $N_p$ increases, which yields $V(\hat{d}) = 0.984$ in Scenario 1 ($V(d^{opt}) = 0.999$) and $V(\hat{d}) = 1.238$ in Scenario 2 ($\max_{d \in \Pi_1} V(d) = 1.251$) when $N_p = 1000$. Finally, the coverage probabilities of the value estimator obtained by CODA-HE are close to the nominal level, which support the theoretical results in Theorem 4.4.3 for heterogeneous baseline covariates.

### 4.6 Real Data Analysis

We illustrate the proposed method by application to data sources from the MIMIC-III clinical database (Goldberger et al. 2000; Johnson et al. 2016; Biseda et al. 2020) and the eICU collaborative research database (Goldberger et al. 2000; Pollard et al. 2018). We consider $r = 11$ common baseline covariates in both samples after dropping variables with high missing rates: age (years), gender (0=female, 1=male), admission weights (kg), admission temperature (Celsius), Glasgow Coma Score (0-15), sodium amount (meq/L), glucose amount (mg/dL), blood urea nitrogen amount (mg/dL), creatinine amount (mg/dL), white blood cell count (E9/L), and total input amount (mL). Here, the treatment is coded as 1 if receiving the vasopressor, and 0 if receiving other medical supervisions such as IV fluid resuscitation. Intermediate outcomes include the total urine output (mL) and the cumulated balance (mL) of metabolism for both samples. The outcome of interest ($Y_p$) is 0 if a patient died due to sepsis and 1 if a patient is still alive, observed only in the primary sample. This accords with the definition of $Y_p$ such that a larger $Y_p$ is better. By deleting the abnormal values in two datasets, we form the primary sample of interest consisting of $N_p = 10746$ subjects, where 2242 patients were treated with the vasopressor while the rest 8504 subjects were treated with other medical supervisions. The auxiliary sample consists of $N_I = 7402$ subjects without the information of the outcome of interest, among which 2005 patients were treated with the vasopressor, while the rest were treated with other medical supervisions.

We illustrate the shared baseline variables in the MIMIC-III data and the eICU data in Fig. 4.1. It can be seen from Fig. 4.1 that there exist multiple variables that have a distinct pattern in each sample, including Glasgow Coma Score, white blood cell count, glucose amount, and blood urea nitrogen amount. This shows some degree of heterogeneity in these two samples. To check the reasonability of assumption (A4), we fit a deep neural network of two intermediate outcomes on baseline covariates and the treatment based on each sample, respectively. For each treatment-covariates pair $(x, a)$ in the support of the two
samples, we estimate $E[M_p | X_p = x, A_p = a]$ based on the fitted deep neural network from the primary sample and $E[M_U | X_U = x, A_U = a]$ based on the fitted deep neural network from the auxiliary sample. The relative mean square error of the difference between the fitted conditional means of two samples over the set of $(x, a)$ is 5.55\% for the total output and 22.60\% for the cumulated balance. This indicates the conditional mean estimators for intermediate outcomes in the two samples are close. Next, we apply CODA on the two samples in comparison to the ODR from the primary sample only. Here, we consider the CODA method for homogenous covariates and heterogeneous covariates, respectively. All the decision rules are searched within the class of decision trees, using a similar procedure introduced in Section 4.5.1. We consider two different sample sizes of the primary sample as $N_p \in \{5000, 10746\}$. The results of the estimated value $\hat{V}$, the estimated standard error $\hat{\sigma}$, the improved efficiency as the relative reduction in the estimated standard error of the CODA value estimator with respect to that of the ODR value estimator, and the number of the assignment to each treatment, are summarized in Table 4.5. In addition, we report the assignment matching rate between CODA and the ODR method based on the primary sample solely in the last row of Table 4.5.

Based on Table 4.5, CODA performs reasonably better than the original ODR under each different $N_p$. Specifically, with the full primary dataset ($N_p = 10746$), the proposed CODA-HO achieves a value of 0.203 with a smaller standard error of 0.0065, comparing to the value under the ODR as 0.192 with a standard error of 0.0068. The efficiency is improved by 4.4\% owing to the CODA-HO. The rate of making the same decision between these two rules is 86.5\%. The CODA-HO assigns 5853 patients to treatment 1 and 4893 patients to the control, which is consistent with the competitive nature of these two treatments. On the other hand, the proposed CODA-HE achieves a value of 0.200 with a slightly smaller standard error of 0.0067, in contrast to the ODR based on the primary sample solely. The

Figure 4.1: The box-plots for the shared baseline variables in the MIMIC-III data and the eICU data.
Table 4.5: The real data analysis under the proposed CODA method and the ODR method based on the primary sample solely.

| Sample Size | $N_p = 5000$ | | | $N_p = 10746$ | | |
|-------------|-------------|----------------|-------------|----------------|----------------|
| Method      | CODA-HO     | CODA-HE        | ODR         | CODA-HO        | CODA-HE        | ODR         |
| Estimated $\hat{V}(\cdot)$ | 0.204 | 0.194 | 0.184 | 0.203 | 0.200 | 0.192 |
| Estimated $\hat{\sigma}$ | 0.0090 | 0.0094 | 0.0097 | 0.0065 | 0.0067 | 0.0068 |
| Improved Efficiency | 7.2% | 3.1% | / | 4.4% | 1.5% | / |
| Treatment 0 | 2967 | 2805 | 2671 | 5853 | 5665 | 5477 |
| Treatment 1 | 2033 | 2195 | 2329 | 4893 | 5081 | 5269 |
| Matching Rate | 87.5% | 93.7% | / | 86.5% | 98.3% | / |

Efficiency is improved by 1.5% based on the CODA-HE. The decision rule under CODA-HE is closer to the ODR. In addition, the proposed CODAs could achieve a greater improvement in efficiency when the sample size is smaller, as 7.2% for CODA-HO and 3.1% for CODA-HE under $N_p = 5000$. These findings are consistent with what we have observed in simulations, which demonstrate that the proposed CODA method offers a statistically innovative and practically pragmatic tool for a more efficient optimal treatment decision making by integrating multiple data sources from heterogeneous studies with the limited outcome.

4.7 Discussions

It is noted that the magnitude of efficiency gain of CODA depends on $\rho(d)$ or $\rho_R(d)$, the correlation between the value estimator of the outcome of interest and the value difference estimator of intermediate outcomes in the two samples. Specifically, for the case with homogeneous baseline covariates, we have

$$
\rho(d) = E\left(\frac{1}{[A_p \pi d(X_p) + (1-A_p)(1-\pi d(X_p))]^2} \times [M - \theta + ]\{X_p, d(X_p)\} \times (A_p \pi d(X_p) + (1-A_p)(1-\pi d(X_p))^2\}
\right)
$$

while for the case with heterogeneous baseline covariates, we have

$$
\rho_R(d) = E\left(\frac{1}{[A_p \pi d(X_p) + (1-A_p)(1-\pi d(X_p))] \times [A_p \pi d(X_p) + (1-A_p)(1-\pi d(X_p))]}
\right)
$$
In general, $\rho(d)$ tends to be larger than $\rho_R(d)$ due to the second summation term in $\rho(d)$ in (4.10), which partly explains why CODA-HO has relatively larger efficiency gain than CODA-HE over the ODR obtained using the primary sample solely as observed in both simulations and the real data application.

There are several possible extensions we may consider in future work. First, we only consider two treatment options in this chapter, while in applications it is common to have more than two options for decision making. Thus, a more general method with multiple treatments or even continuous decisions is desirable. Second, we can extend the proposed CODA method to dynamic treatment decision making, where each subject successively receives a treatment followed by intermediate outcomes, however, the primary outcome of interest can be observed in the primary sample only. Third, we only consider the setting where two samples share the same set of baseline covariates so that the ignorability assumption holds in both samples. In practice, different samples from heterogeneous studies may not have exactly the same set of baseline covariates. To be specific, let $X_P = [X_P^{(1)}, \cdots, X_P^{(r_1)}]^\top$ denote $r_1$-dimensional individual’s baseline covariates in the primary sample, and let $X_U = [X_U^{(1)}, \cdots, X_U^{(r_2)}]^\top$ denote $r_2$-dimensional individual’s baseline covariates in the auxiliary sample. The ignorability assumption holds for its own set of covariates in each sample. Suppose two samples share a same subset of baseline covariates with dimension $r_3 \leq \min(r_1, r_2)$, denoted as $X_C$. Suppose that $X_C$ has the same joint distribution in two samples and the comparable intermediate outcomes assumption holds for this common set of covariates $X_C$. Then, we can modify the proposed calibrated value estimator by calibrating only with this common set of baseline covariates in two samples, but maintain searching the ODR based on whole available baseline covariates in the primary sample. We leave it for future research.
5.1 Introduction

We focus on addressing the challenge of developing the optimal IDR when the long-term outcome cannot be observed in the experimental sample. Although the long-term outcome may not be observed in the experimental sample, we could instead obtain some intermediate outcomes (also known as surrogacies or proximal outcomes, $M$) that are highly related to the long-term outcome after the treatment was given. For instance, the CD4 and CD8 counts recorded after a treatment is assigned, have a strong correlation with the healthy of the immune system, and thus can be viewed as intermediate outcomes. A natural question is whether an optimal IDR to maximize the expected long-term outcome can be estimated based on the experimental sample (that consists of $\{X, A, M\}$) only. The answer is generally no mainly for two reasons. First, it is common and usually necessary
to have multiple intermediate outcomes to characterize the effects of treatment on the long-term outcome. However, when there are multiple intermediate outcomes, it is hard to determine which intermediate outcome or what combination of intermediate outcomes will lead to the best IDR for the long-term outcome. Second, to derive the optimal IDR that maximizes the expected long-term outcome of interest based on the experimental sample, we need to know the relationship between the long-term outcome, intermediate outcomes and baseline covariates, which is generally not practical.

In this thesis, we propose using an auxiliary data source, namely the auxiliary sample, to recover the missing long-term outcome of interest in the experimental sample, based on the rich information of baseline covariates and intermediate outcomes. Auxiliary data, such as electronic medical records or administrative records, are now widely accessible. These data usually contain rich information for covariates, intermediate outcomes, and the long-term outcome of interest. However, since they are generally not collected for studying treatment effects, treatment information may not be available in auxiliary data. In particular, in this work, we consider the situation that an auxiliary data consisting of \( \{X, M, Y\} \) is available, where \( Y \) is the long-term outcome of interest. Note it is also impossible to derive ODR based on such auxiliary sample due to missing treatments.

Our considered estimation of the optimal IDR naturally falls in the framework of semi-supervised learning. A large number of semi-supervised learning methods have been proposed for the regression or classification problems (Zhu 2005; Chen et al. 2008; Chapelle et al. 2009; Chakrabortty et al. 2018). Recently, Athey et al. (2019) studied the estimation of the average treatment effect under the framework of combining the experimental data with the auxiliary data. They proposed to use the surrogate index and clarified the comparability and surrogacy assumptions, which allowed them to impute the missing outcomes in the experimental data based on the regression model learned from the auxiliary data using baseline covariates and intermediate outcomes. However, as far as we know, no work has been done for estimating the optimal IDR in such a semi-supervised setting.

Our work contributes to the following folds. First, to the best of our knowledge, this is the first work on estimating the heterogeneous treatment effect and developing the optimal decision making for the long-term outcome that cannot be observed in an experiment, by leveraging the idea from semi-supervised learning. Methodologically, we propose an auGmented inverse propensity weighted Experimental and Auxiliary sample-based decision Rule, named GEAR. This rule maximizes the augmented inverse propensity weighted estimator (AIPW, also known as the doubly-robust estimator) of the value function over a class of interested decision rules using the experimental sample, with the primary out-
come being imputed based on the auxiliary sample. Theoretically, we show that the AIPW estimator under the proposed GEAR is consistent and derive its corresponding asymptotic distribution under certain conditions. A confidence interval (CI) for the estimated value is provided.

The rest of this chapter is organized as follows. We introduce the statistical framework for estimating the optimal treatment decision rule using the experimental sample and the auxiliary sample, and associated assumptions in Section 5.2. In Section 5.3, we propose our GEAR method and establish consistency and asymptotic distributions of the estimated value functions under the proposed GEAR. Extensive simulations are conducted to demonstrate the empirical validity of the proposed method in Section 5.4, followed by an application to ACTG 175 data in Section 5.5. We conclude this chapter with a discussion in Section 5.6. The technical proofs and sensitivity studies under model assumption violation are given in the appendix.

5.2 Statistical Framework

5.2.1 Experimental Sample and Auxiliary Sample

Suppose there is an experimental sample of interest \( E \). Let \( X_E \) denote \( r \)-dimensional individual's baseline covariates with the support \( X_E \in \mathbb{R}^r \), and \( A_E \in \{0, 1\} \) denote the treatment an individual receives. The long-term outcome of interest \( Y_E \) with support \( Y_E \in \mathbb{R} \) cannot be observed, instead we only obtain the \( s \)-dimensional intermediate outcomes \( M_E \) with support \( M_E \in \mathbb{R}^s \) after a treatment \( A_E \) is assigned. Denote \( N_E \) as the sample size for the experimental sample, which consists of \( \{E_i = (X_{E,i}, A_{E,i}, M_{E,i}), i = 1, \ldots, N_E\} \) independent and identically distributed (I.I.D.) across \( i \).

To recover the missing long-term outcome of interest in the experimental sample, we include an auxiliary sample, \( U \), which contains the individual's baseline covariates \( X_U \), intermediate outcomes \( M_U \), and the observed long-term outcome of interest \( Y_U \), with support \( X_U, M_U, Y_U \) respectively. However, treatment information is not available in the auxiliary sample. Let \( N_U \) denote the sample size for the I.I.D. auxiliary sample that includes \( \{U_i = (X_{U,i}, M_{U,i}, Y_{U,i}), i = 1, \ldots, N_U\} \).

We use \( R = \{E, U\} \) to indicate the missingness and identification of each sample, where \( R = E \) implies the experimental sample with missing long-term primary outcome and \( R = U \) means the auxiliary sample with missing treatment information. Thus, these two samples can also be rewritten as one joint sample \( \{(X_i, R_i, A_i\mathbb{I}_{R_i = E}, M_i, Y_i\mathbb{I}_{R_i = U}), i = 1, \ldots, N_E + N_U\} \).
where $\mathbb{I}(\cdot)$ is an indicator function.

### 5.2.2 Assumptions

In this subsection, we make five key assumptions in order to introduce the ODR. For the experimental sample, define the potential outcomes $Y_E^*(0)$ and $Y_E^*(1)$ as the long-term outcome that would be observed after an individual receiving treatment 0 or 1, respectively. Let the propensity score as the conditional probability of receiving treatment 1 in the experimental sample, i.e. $\pi(x) = P r_E(A_{E,i} = 1 | X_{E,i} = x)$. As standard in causal inference by Rubin (1978), we assume:


(A2). No Unmeasured Confounders Assumption: $\{Y_E^*(0), Y_E^*(1)\} \perp \perp A_E | X_E$.

(A3). $0 < \pi(x) < 1$ for all $x \in X_E$.

To impute the missing long-term outcome in the experimental sample with the assistance of the auxiliary sample, we introduce the following two assumptions, the comparability assumption and the surrogacy assumption.

First, the comparability assumption states that the population distribution of the long-term outcome of interest $Y$ is independent of whether belonging to the experimental sample or the auxiliary sample, given the information of population baseline covariates $X$ and population intermediate outcomes $M$ as follows.

(A4). Comparability Assumption: $Y \perp \perp R | X, M$.

Here, (A4) is also known as 'conditional independence assumption' made in Chen et al. (2008), and has an equivalent expression as $Y_E | \{M_E, X_E\} \sim Y_U | \{M_U, X_U\}$ proposed in Athey et al. (2019). When (A4) holds, we have a direct conclusion of the equality of the conditional mean outcome given baseline covariates and intermediate outcomes in each sample, stated in the following corollary.

**Corollary 5.2.1. (Equal Conditional Mean) Under (A4),**

$$E[Y_E | M_E = m, X_E = x] = E[Y_U | M_U = m, X_U = x].$$ (5.1)

**Remark 5.2.1.** It is shown in Section 5.3 that (A4) can be relaxed to Equation (5.1) for deriving the proposed method.

We further define the missing at random (MAR) assumption in the joint sample as: $\{Y, A\} \perp \perp R | X, M$; and give the following corollary to show the relationship between (A4) and the MAR assumption.
Figure 5.1: A direct acyclic graph illustrating assumptions (A2), (A4), and (A5) in the joint sample. White nodes represent observed variables, and grey nodes are variables with missing values.

**Corollary 5.2.2.** (MAR Assumption)

\[
\{Y, A\} \perp R \mid X, M \quad \longrightarrow \quad Y \perp R \mid X, M.
\]

**Remark 5.2.2.** Corollary 5.2.2 is a direct result of joint independence implying marginal independence. Though (A4) is untestable due to the missing long-term outcome in the experimental sample, one can believe (A4) holds if there exists strong evidence about the reasonability of the MAR assumption in the joint sample.

Second, the surrogacy assumption states that the long-term outcome of interest in the experimental sample is independent of the treatment conditional on a set of baseline covariates and intermediate outcomes as below.

(A5). Surrogacy Assumption: \( Y_E \perp A_E \mid X_E, M_E \).

**Remark 5.2.3.** The above assumption is also used in Athey et al. (2019). The validation of the surrogacy assumption relies on the ‘richness’ of intermediate outcomes that are highly related to the long-term outcome of interest. Similarly, it is infeasible to check the surrogacy assumption due to the missing long-term outcome in the experimental sample.

We illustrate the statistical framework of the joint sample under above assumptions by a direct acyclic graph in Figure 5.1. Graphically, \( A \) and \( Y \) have no common parents except for \( X \), encoding (A2); \( R \) and \( Y \) have two common parents, \( X \) and \( M \), encoding (A4); when fixing \( X \) and \( M \), \( A \) and \( Y \) are independent, encoding (A5).
5.2.3 Value Function and Optimal Decision Rule

A decision rule is a deterministic function \( d(\cdot) \) that maps \( X_E \) to \{0, 1\}. Define the potential outcome of interest under \( d(\cdot) \) as \( Y_E^*(d) = Y_E^*(0)\{1 - d(X_E)\} + Y_E^*(1)d(X_E) \), which would be observed if a randomly chosen individual from the experimental sample had received a treatment according to \( d(\cdot) \), where we suppress the dependence of \( Y_E^*(d) \) on \( X_E \). We then define the value function under \( d(\cdot) \) as the expectation of the potential outcome of interest over the experimental sample as

\[
V(d) = \mathbb{E}\{Y_E^*(d)\} = \mathbb{E}[Y_E^*(0)\{1 - d(X_E)\} + Y_E^*(1)d(X_E)].
\]

As a result, we have the optimal treatment decision rule (ODR) of interest defined to maximize the value function over the experimental sample among a class of decision rules of interest as \( d^{opt}(\cdot) = \arg\min_{d(\cdot)} V(d) \). Suppose the decision rule \( d(\cdot) \) relies on a model parameter \( \beta \), denoted as \( d(\cdot) \equiv d(\cdot; \beta) \). We use a shorthand to write \( V(d) \) as \( V(\beta) \), and define \( \beta_0 = \arg\min_{\beta} V(\beta) \). Thus, the value function under the true ODR \( d(\cdot; \beta_0) \) is defined as \( V(\beta_0) \).

5.3 Proposed Method

In this section, we detail the proposed method by constructing the AIPW value estimator for the long-term outcome based on two samples. Implementation details are provided to find the ODR. The consistency and asymptotical distribution of the value estimator under our proposed method are presented, followed by its confidence interval. All the proofs are provided in Section B of the appendix.

5.3.1 Proposed Estimator for Long-Term Outcome

To overcome the difficulty of estimating the value function due to the missing long-term outcome of interest in the experimental sample, one intuitive way is to impute the missing outcome \( Y_E \) with its conditional mean outcome given baseline covariates and intermediate outcomes (total common information available in both samples).

Denote \( \mu_E(m, x) \equiv \mathbb{E}[Y_E|M_E = m, X_E = x] \), and \( \mu_U(m, x) \equiv \mathbb{E}[Y_U|M_U = m, X_U = x] \). Under Corollary 5.2.1, we have \( \mu_E(m, x) = \mu_U(m, x) \). Here, \( \mu_E(m, x) \) is inestimable because of the missing long-term outcome. We instead use \( \mu_U(M_E, X_E) \) to impute the missing \( Y_E \).
and give the following lemma as a middle step to construct the AIPW value estimator for the long-term outcome.

**Lemma 5.3.1.** Under (A1)-(A5), given \( d(\cdot; \beta) \), we have

\[
V(\beta) = E \left[ \frac{1}{A_E \pi(X_E) + (1 - A_E)\{1 - \pi(X_E)\}} \right].
\]

Next, we propose the AIPW estimator of the value function for the long-term outcome in the experimental sample. To address the difficulty of forming the augmented term when the long-term outcome of interest cannot be observed, we show that augmenting on the missing long-term outcome is equivalent to augmenting on the imputed conditional mean outcome of interest \( \mu_U(M_E, X_E) \), by the following lemma.

**Lemma 5.3.2.** Under (A1)-(A5), given \( d(\cdot; \beta) \), we have

\[
E_{Y_E|X_E} \{ Y_E | A_E = d(X_E; \beta), X_E \} = E_{M_E|X_E} \{ \mu_U(M_E, X_E) | A_E = d(X_E; \beta), X_E \},
\]

where \( E_{A|B} \) means taking expectation with respect to the conditional distribution of \( A \) given \( B \).

According to Lemma 5.3.1 and Lemma 5.3.2, given a decision rule \( d(\cdot; \beta) \), the value function \( V(\beta) \) can be consistently estimated through

\[
V^*_{n_{AIP}}(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \left[ \nu_i + \frac{1}{A_E \pi(X_E) + (1 - A_E)\{1 - \pi(X_E)\}} \right],
\]

where \( \nu_i \equiv E\{\mu_U(M_{E,i}, X_{E,i}) | A_{E,i} = d(X_{E,i}; \beta), X_{E,i}\} \) presents the augmented term. Here, the propensity score \( \pi \) can be estimated in the experimental sample, denoted as \( \hat{\pi} \), and the conditional mean \( \mu_U \) can be estimated in the auxiliary sample, denoted as \( \hat{\mu}_U \). Then, by replacing the implicit functions in \( V^*_{n_{AIP}}(\beta) \), it is straightforward to give the AIPW estimator of the value function \( V(\beta) \) as

\[
\hat{V}_{AIP}(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \left[ \hat{\nu}_i + \frac{1}{A_E \hat{\pi}(X_E) + (1 - A_E)\{1 - \hat{\pi}(X_E)\}} \right],
\]

where \( \hat{\nu}_i \equiv E\{\hat{\mu}_U(M_{E,i}, X_{E,i}) | A_{E,i} = d(X_{E,i}; \beta), X_{E,i}\} \) is the estimator for \( \nu_i \). We define \( \hat{\beta}^G = \arg\max_\beta \hat{V}_{AIP}(\beta) \), and then propose the augmented inverse propensity weighted Experimental and Auxiliary sample-based decision Rule (GEAR) for the long-term outcome of
interest as \(d(X; \hat{\beta}^G)\) with the corresponding estimated value function as \(\hat{V}_{AIP}(\hat{\beta}^G)\).

### 5.3.2 Implementation Details

**Class of decision rules:** The GEAR can be searched within a pre-specified class of decision rules. Popular classes include generalized linear rules, fixed depth decision trees, threshold rules, and so on (Zhang et al. 2012b; Athey and Wager 2017; Rai 2018). In this chapter, we focus on the class of generalized linear rules. Specifically, suppose the decision rule takes a form as \(d(X_E; \beta) \equiv \mathbb{I}[g(X_E)^T \beta > 0]\), where \(g(\cdot)\) is an unknown function. We use \(\phi_X(\cdot)\) to denote a set of basis functions of \(X_E\) with length \(v\), which are “rich” enough to approximate the underlying function \(g(\cdot)\). Thus, the GEAR is found within a class of \(\mathbb{I}\{\phi_X(X_E)^T \beta > 0\}\). For notational simplicity, we include 1 in \(\phi_X(\cdot)\) so that \(\beta \in \mathbb{R}^{v+1}\). With subject to \(||\beta||_2 = 1\) for identifiability purpose, the maximizer for \(\hat{V}_{AIP}(\beta)\) can be solved using any global optimization algorithm. In our implementation, we apply the heuristic algorithm to search for the GEAR.

**Estimation models:** The conditional mean of the long-term outcome \(\mu_U(m, x)\) can be estimated through any parametric or nonparametric model. In practice, we assume \(\mu_U(m, x)\) can be determined by a flexible basis function of baseline covariates and intermediate outcomes, to fully capture the underlying true model. Similarly, one can use a flexible basis function of baseline covariates and the treatment to model the augmented term as well as the propensity score function. Note that any machine learning tools such as Random Forest or Deep Learning can be applied to model terms in the proposed AIPW estimator. Our theoretical results still hold under these nonparametric models as long as the regressors have desired convergence rates (see results established in Wager and Athey (2018); Farrell et al. (2018)).

**Estimation of the augmented term:** To estimate the augmented term \(\nu_i\), we need three steps as follows. First, we model \(\mu_U(m, x)\) through the auxiliary sample \(\{X_U, M_U, Y_U\}\) as \(\hat{\mu}_U(m, x)\); second, we plug \(\{M_E, X_E\}\) of the experimental sample into \(\hat{\mu}_U(m, x)\) and get \(\hat{\mu}_U(M_E, X_E)\) as the conditional mean outcome of interest to impute the missing \(Y_E\); at last, we fit \(\hat{\mu}_U(M_E, X_E)\) on \(\{A_E, X_E\}\) in the experimental sample, and get \(\hat{\nu}_i\).

### 5.3.3 Theoretical Properties

We next show the consistency and asymptotic normality of our proposed AIPW estimator. Its asymptotic variance can be decomposed into two parts, corresponding to the estimation
variances from two independent samples. As mentioned in Section 5.3.2, our AIPW estimator can handle various machine learning or parametric estimators as long as regressors have desired convergence rates. To derive an explicit variance form, we next focus on parametric models.

We posit parametric models for $\pi(x) \equiv \pi(x; \gamma)$ and $\mu_U(m, x) \equiv \mu_U(m, x; \lambda)$ with true model parameters $\gamma$ and $\lambda$. Let $\phi_X(X)$ and $\phi_M(M)$ to represent appropriate basis functions for $X$ and $M$, respectively. Without loss of generality, we posit basis model for the augmented term such that $E[\mu_U(m, x; \lambda) | A = 0, X = x] \equiv \phi_X(x) \theta_0$, and $E[\mu_U(m, x; \lambda) | A = 1, X = x] \equiv \phi_X(x) \theta_1$ with true model parameters $\theta_0$ and $\theta_1$. The following conditions are needed to derive our theoretical results:

(A6). Suppose the density of covariates $f_X(x)$ is bounded away from 0 and $\infty$ and is twice continuously differentiable with bounded derivatives.

(A7). Both $\pi(x; \gamma)$ and $\mu_U(m, x; \lambda)$ are smooth bounded functions, with their first derivatives exist and bounded.

(A8). Model for $\mu_U(m, x; \lambda)$ is correctly specified.

(A9). Denote $t = \sqrt{\frac{N_0}{N_0}}$ and assume $0 < t < +\infty$.

(A10). The true value function $V(\beta)$ is twice continuously differentiable at a neighborhood of $\beta_0$.

(A11). Either the model of the propensity score or the model of the augmented term is correctly specified.

Here, (A6) and (A10) are commonly imposed to establish the inference for value search methods (Zhang et al. 2012b; Wang et al. 2018). (A7) is assumed for desired convergence rates of $\hat{\pi}$ and $\hat{\mu}_U$. To apply machine learning tools, similar assumption is required (see more details in Wager and Athey (2018); Farrell et al. (2018)). From (A8), we can replace the missing long-term outcome with its imputation, and thus the consistency holds. Evaluations are provided in Section 5.4.2 to examine the proposed method when (A8) is violated. (A9) states that the sizes of two samples are comparable, which prevents the asymptotic variance from blowing up when combining two samples in semi-supervised learning (Chen et al. 2008; Chakrabortty et al. 2018). (A11) is included to establish the doubly robustness of the value estimator, which is commonly used in the literature of doubly robust estimator (Dudík et al. 2011; Zhang et al. 2012b, 2013c).

The following theorem gives the consistency of our AIPW estimator of the value function to the true value function.
Theorem 5.3.1. (Consistency) Under (A1)-(A9) and (A11),

\[ \hat{V}_{AIP}(\beta) = V(\beta) + o_p(1), \quad \forall \beta. \]

Remark 5.3.1. When the model for \( \mu_{U}(m, x) \) is correctly specified, our AIPW estimator is doubly robust given either the model of the propensity score or the model of the augmented term is correct. To prove the theorem, we establish the theoretical results with their proofs for the inverse propensity-score weighted estimator as a middle step. See more details in Section A of the appendix.

To establish the asymptotic normality of \( \hat{V}_{AIP}(\hat{\beta}^G) \), we first show the estimator \( \hat{\beta}^G \) has a cubic rate towards the true \( \beta_0 \).

Lemma 5.3.3. Under (A1)-(A11), we have

\[ N_{1/3}^E \| \hat{\beta}^G - \beta_0 \|_2 = O_p(1), \quad \text{(5.2)} \]

where \( \| \cdot \|_2 \) is the L_2 norm, and \( O_p(1) \) means the random variable is stochastically bounded.

Based on Lemma 5.3.3, we next give the asymptotic normality of \( \sqrt{N_E} \{ \hat{V}_{AIP}(\hat{\beta}^G) - V(\beta_0) \} \) in the following theorem.

Theorem 5.3.2. (Asymptotic Distribution) Under (A1)-(A11),

\[ \sqrt{N_E} \{ \hat{V}_{AIP}(\hat{\beta}^G) - V(\beta_0) \} \xrightarrow{q} N(0, \sigma_{AIP}^2), \quad \text{(5.3)} \]

where \( \sigma_{AIP}^2 = t^2 \sigma_U^2 + \sigma_E^2, \sigma_U^2 = E[\xi^{(U)}_i]^2, \) and \( \sigma_E^2 = E[\xi^{(E)}_i]^2 \). Here, \( \xi^{(E)}_i \) and \( \xi^{(U)}_i \) are the I.I.D. terms in the experimental sample and auxiliary sample, respectively.

Remark 5.3.2. From Theorem 5.3.2, the asymptotic variance of the AIPW estimator has an additive form that consists of the estimation error from each sample. Proportion of these two estimation variances is controlled by the sample ratio. In reality, \( N_U \) is usually larger than \( N_E \). When \( N_U/N_E \to \infty \), we have \( t \to 0 \), and thus the estimation error from auxiliary sample can be ignored. Our result under this special case is supported by Chakrabortty et al. (2018) where they considered \( N_U/N_E \to \infty \) for a regression problem.

Next, we give explicit form of \( \xi^{(E)}_i \) and \( \xi^{(U)}_i \) from the proof of Theorem 5.3.2 to estimate
\( \sigma_{AIP} \). Denote \( \pi(x;\gamma) \equiv \partial \pi(x;\gamma)/\partial \gamma \) and \( \mu_U(m, x; \lambda) \equiv \partial \mu_U(m, x; \lambda)/\partial \lambda \). Let

\[
H_1 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \phi_X(X_{E,i}) \pi(X_{E,i}; \gamma) \top,
\]

\[
H_2 \equiv \lim_{N_U \to \infty} \frac{1}{N_U} \sum_{i=1}^{N_U} \left[ \phi_X(X_{U,i}) \right] \mu_U(M_{U,i}, X_{U,i}; \lambda) \top,
\]

\[
H_3 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} (1 - A_{E,i}) \phi_X(X_{E,i}) \phi_X(X_{E,i}) \top,
\]

\[
H_4 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} A_{E,i} \phi_X(X_{E,i}) \phi_X(X_{E,i}) \top,
\]

\[
G_1 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} r_i (1 - 2A_{E,i}) \pi(X_{E,i}; \gamma) \mu_U(M_{E,i}, X_{E,i}; \lambda) /
\]

\[
s_i^2,
\]

\[
G_2 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{r_i}{s_i} \mu_U(M_{E,i}, X_{E,i}; \lambda),
\]

\[
G_3 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{r_i}{s_i} q_i (1 - 2A_{E,i}) \pi(X_{E,i}; \gamma) /
\]

\[
s_i^2,
\]

\[
G_4 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \left[ 1 - \frac{r_i}{s_i} \right] \phi_X(X_{E,i}) \{1 - d(X_{E,i}; \beta_0)\},
\]

\[
G_5 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \left[ 1 - \frac{r_i}{s_i} \right] \phi_X(X_{E,i}) d(X_{E,i}; \beta_0),
\]

where \( r_i \equiv 1\{A_{E,i} = d(X_{E,i}; \beta_0)\}, s_i \equiv A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})\{1 - \pi(X_{E,i}; \gamma)\}, \) and \( q_i \equiv \phi_X(X_{E,i}) \top \theta_0 + \phi_X(X_{E,i}) \top (\theta_1 - \theta_0) d(X_{E,i}; \beta_0) \). Then, the I.I.D. term in the experimental sample is

\[
\xi^{(E)}_i \equiv \frac{r_i [\mu_U(M_{E,i}, X_{E,i}; \lambda) - \nu^*_i]}{s_i} + \nu^*_i - V(\beta_0)
\]

\[
+ (G_1^\top + G_3^\top) H_1^{-1} \phi_X(X_{E,i}) \{A_{E,i} - \pi(X_{E,i}; \gamma)\}
\]

\[
+ G_5^\top H_4^{-1} \phi_X(X_{E,i}) A_{E,i} \{\mu_U(M_{E,i}, X_{E,i}; \lambda) - \phi_X(X_{E,i}) \top \theta_1\}
\]

\[
+ G_4^\top H_3^{-1} \phi_X(X_{E,i}) \{1 - A_{E,i}\} \{\mu_U(M_{E,i}, X_{E,i}; \lambda) - \phi_X(X_{E,i}) \top \theta_0\},
\]

for \( \nu^*_i \equiv E\{\mu_U(M_{E,i}, X_{E,i}; \lambda) | A_{E,i} = d(X_{E,i}; \beta_0), X_{E,i}\} \). And

\[
\xi^{(U)}_i \equiv G_2^\top H_2^{-1} \left[ \phi_X(X_{U,i}) \right] \{Y_{U,i} - \mu_U(M_{U,i}, X_{U,i}; \lambda)\},
\]

92
corresponds to the I.I.D. term in the auxiliary sample.

By plugging the estimations into the pre-specified models, we could obtain the estimated \( \hat{\xi}_i^{(E)} \) and \( \hat{\xi}_i^{(U)} \). Then the variance \( \sigma^2_E \) and \( \sigma^2_U \) can be consistently estimated by \( \hat{\sigma}^2_E = \frac{1}{N_E} \sum_{i=1}^{N_E} \{ \hat{\xi}_i^{(E)} \}^2 \) and \( \hat{\sigma}^2_U = \frac{1}{N_U} \sum_{i=1}^{N_U} \{ \hat{\xi}_i^{(U)} \}^2 \), respectively. Thus, we can estimate \( \sigma_{Alp} \) through

\[
\hat{\sigma}_{Alp} = \sqrt{t^2 \hat{\sigma}^2_U + \hat{\sigma}^2_E},
\]

based on Theorem 5.3.2. Therefore, a two-sided \( 1 - \alpha \) confidence interval (CI) for \( V(\beta_0) \) under the GEAR is

\[
\left[ \hat{V}_{Alp}(\hat{\beta}^G) - \frac{z_{\alpha/2} \hat{\sigma}_{Alp}}{\sqrt{N_E}}, \hat{V}_{Alp}(\hat{\beta}^G) + \frac{z_{\alpha/2} \hat{\sigma}_{Alp}}{\sqrt{N_E}} \right],
\]

where \( z_{\alpha/2} \) denotes the upper \( \alpha/2 \)-th quantile of a standard normal distribution.

## 5.4 Simulation Studies

In this section, we evaluate the proposed method when the model of the conditional mean of the long-term outcome correctly specified and misspecified separately in the following two subsections. Additional sensitivity studies when assumptions are violated are provided in Section C of the appendix.

### 5.4.1 Evaluation under Correctly Specified Model

Simulated data, including baseline covariates \( X = [X^{(1)}, X^{(2)}, \cdots, X^{(r)}]^T \), the treatment \( A \), intermediate outcomes \( M = [M^{(1)}, M^{(2)}, \cdots, M^{(s)}]^T \), and the long-term outcome \( Y \), are generated from the following model:

\[
X^{(1)}, X^{(2)}, \cdots, X^{(r)} \sim iid \text{ Uniform}[−1, 1], \quad A \sim \text{Bernoulli}(0.5),
M = H^M(X) + AC^M(X) + \epsilon^M, \quad Y = H^Y(X) + C^Y(X, M) + \epsilon^Y,
\]

where \( \epsilon^M \) and \( \epsilon^Y \) are random errors following \( N(0, 0.5) \). Here, \( A \) in the auxiliary sample is used only for generating intermediate outcomes such that the comparability assumption is satisfied. Note that \( Y \) is generated for the auxiliary sample only. Given \( X \) and \( M \), we can see \( Y \) is independent of \( A \), which indicates the surrogacy assumption.
Set \( r = 4 \) and \( s = 2 \). We consider following two scenarios with different \( H^M(\cdot) \), \( C^M(\cdot) \), \( H^Y(\cdot) \), and \( C^Y(\cdot) \).

**S1:**

\[
\begin{align*}
H^M(X) &= \begin{bmatrix} X^{(3)} \\ X^{(1)} \end{bmatrix}, \quad C^M(X) = \begin{bmatrix} 4\{X^{(1)}-X^{(2)}\} \\ 4\{X^{(4)}-X^{(3)}\} \end{bmatrix}, \\
H^Y(X) &= -1 + X^{(2)} + X^{(4)}, \quad C^Y(X, M) = M^{(1)} + M^{(2)}.
\end{align*}
\]

**S2:**

\[
\begin{align*}
H^M(X) &= \begin{bmatrix} (X^{(1)})^2 X^{(3)} + \sin\{X^{(4)}\} \\ (X^{(1)})^3 - \{X^{(2)} - X^{(4)}\}^2 \end{bmatrix}, \\
C^M(X) &= \begin{bmatrix} 4\{X^{(1)}-X^{(2)}\} \\ 4\{X^{(4)}-X^{(3)}\} \end{bmatrix}, \\
H^Y(X) &= -1 + X^{(2)} + X^{(4)}, \quad C^Y(X, M) = M^{(1)} + M^{(2)}.
\end{align*}
\]

Under Scenario 1 and 2, we have the parameter of the true ODR as \( \beta_0 = [0, 0.5, -0.5, -0.5, 0.5]^T \) with subject to \( ||\beta_0||_2 = 1 \), which can be easily solved based on the function \( C^M(\cdot) \) that describes the treatment-covariates interaction. The true value \( V(\beta_0) \) can be calculated by Monte Carlo approximations, as listed in Table 5.1. We consider \( N_U = 400 \) for the auxiliary sample and allow \( N_E \) chosen from the set \{200, 400, 800\} in the experimental sample.

To apply the GEAR, we model the conditional mean of the long-term outcome \( \mu_U(m, x) \) and the augmented term \( v_i \) in the auxiliary data via a linear regression. Here, the model of \( \mu_U(m, x) \) is correctly specified by noting that \( Y \) is linear in \( \{X, M\} \) under Scenario 1 and 2. The GEAR is searched within a class of \( d(X_E; \beta) = \mathbb{I}(X_E^\top \beta > 0) \) subjecting to \( ||\beta||_2 = 1 \), through Genetic Algorithm provided in R package *rgenoud*, where we set ‘optim.method’ = ‘Nelder-Mead’, ‘pop.size’ = 3000, ‘domain’=[-10,10], and ‘starting.values’ as a zero vector. Results are summarized in Table 5.1, including the estimated value under the estimated rule \( \hat{V}_{AIP}(\hat{\beta}^G) \) and its standard error \( SE\{\hat{V}_{AIP}\} \), the estimated standard deviation \( \text{E}\{\hat{\sigma}_{AIP}\} \) by Equation (5.4), the value under the estimated rule \( V(\hat{\beta}^G) \) by plugging the GEAR into the true model, the empirical coverage probabilities (CP) for 95\% CI constructed by Equation (5.5), the rate of the correct decision (RCD) made by the GEAR, and the \( L_2 \) loss of \( \hat{\beta}^G (||\hat{\beta}^G - \beta_0||_2) \), aggregated over 500 simulations.

From Table 5.1, it is clear that both the estimated GEAR and its estimated value approach to the true as the sample size \( N_E \) increases in all scenarios. Specifically, our proposed GEAR method achieves \( V(\hat{\beta}^G) = 0.86 \) in Scenario 1 \( (V(\beta_0) = 0.87) \) and \( V(\hat{\beta}^G) = 0.19 \) in Scenario 2 \( (V(\beta_0) = 0.20) \) when \( N_E = 800 \). Notice that the \( \ell_2 \) loss of \( \hat{\beta}^G \) decays at a rate that is
Table 5.1: Empirical results under the GEAR for Scenarios 1 and 2.

<table>
<thead>
<tr>
<th></th>
<th>Scenario 1</th>
<th>Scenario 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_E =$</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>$V(\beta_0)$</td>
<td>0.87</td>
<td>0.20</td>
</tr>
<tr>
<td>$\hat{V}_{AIP}(\hat{\beta}^G)$</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>$SE{\hat{V}_{AIP}}$</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>$E{\hat{\sigma}_{AIP}}$</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>$V(\hat{\beta}^G)$</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>CP (%)</td>
<td>94.6</td>
<td>94.8</td>
</tr>
<tr>
<td>RCD (%)</td>
<td>95.9</td>
<td>96.6</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\hat{\beta}^G - \beta_0</td>
</tr>
</tbody>
</table>

approximately proportional to $N_E^{-1/3}$, which verifies our theoretical findings in Lemma 5.3.3. Moreover, the average rate of the correct decision made by the GEAR increases with $N_E$ increasing. In addition, there are two findings that help to verify Theorem 5.3.2. First, the estimated standard deviation of value function is close to the standard error of the estimated value function, and gets smaller as the sample size $N_E$ increases. Second, the empirical coverage probabilities of the proposed 95% CI approach to the nominal level in all settings. Note that there is no strictly increasing trend of the empirical coverage probabilities due to the fixed sample size $N_U = 400$.

### 5.4.2 Evaluation under Model Misspecification

We consider more general settings to examine the proposed method when the model of $\mu_U(m, x)$ is misspecified. The data is generated from the same model in Section 5.4.1. We fix $H^M(X) = \begin{bmatrix} X^{(3)} \\ X^{(1)} \end{bmatrix}$, $C^M(X) = \begin{bmatrix} 4\{X^{(1)} - X^{(2)}\} \\ 4\{X^{(4)} - X^{(3)}\} \end{bmatrix}$, and set following three scenarios with different
\( H^Y(\cdot) \) and \( C^Y(\cdot) \).

\[
\begin{align*}
S3 : \quad & \begin{cases} 
H^Y(X) = \{X^{(1)} + X^{(3)}\}\{X^{(1)}\}^2 \\
+ \sin\{X^{(4)}\} - \{X^{(2)} - X^{(4)}\}^2,
C^Y(X, M) = M^{(1)} + M^{(2)}. 
\end{cases}
\end{align*}
\]

\[
\begin{align*}
S4 : \quad & \begin{cases} 
H^Y(X) = \{X^{(1)}\}^3 + \{X^{(2)}\}^2 + X^{(3)}, \\
C^Y(X, M) = M^{(1)} + X^{(4)}M^{(2)}. 
\end{cases}
\end{align*}
\]

\[
\begin{align*}
S5 : \quad & \begin{cases} 
H^Y(X) = X^{(2)} - \{X^{(4)}\}^2, \\
C^Y(X, M) = 0.25\{M^{(1)} - X^{(3)}\}^2 + M^{(2)}. 
\end{cases}
\end{align*}
\]

Under Scenario 3, we have the true ODR is still linear while the true ODRs for Scenario 4 and 5 are non-linear due to their \( C^Y(\cdot) \) involving covariates-surrogacy interaction. Table 5.2 lists the true value \( V(\beta_0) \) for each scenario.

We apply the proposed GEAR with the tensor-product B-splines for Scenario 3-5, respectively. Specifically, we first model \( \mu_U(m, x) \) with the tensor-product B-splines of \( \{X_U, M_U\} \) in the auxiliary sample. The degree and knots for the B-splines are selected based on five-fold cross validation to minimize the least square error of the linear regression. Then, we search the GEAR within the class of \( \mathbb{I}\{\phi_X(X_E)^\top \beta > 0\} \), where \( \phi_X(\cdot) \) is the polynomial basis with degree=2. Here, the augmented term is fitted by a linear regression of \( \hat{\mu}_U(M_E, X_E) \) on \( \{A_E, \phi_X(X_E)\} \). We name the above procedure as ‘GEAR-Bspline’. For comparison, we also apply the linear procedure described in Section 5.4.1 as ‘GEAR-linear’ without taking any basis. One may note both procedures model \( \mu_U(m, x) \) incorrectly. Reported in Table 5.2 are the empirical results under GEAR-Bspline and GEAR-linear aggregated over 500 simulations.

It can be seen from Table 5.2 that the GEAR-Bspline procedure performs reasonably better than the linear procedure under non-linear decision rules. Specifically, in Scenario 3 with only the baseline function \( H^Y(\cdot) \) non-linear in \( X \), GEAR-linear performs comparable to GEAR-Bspline, as the linear model can well approximate the non-linear baseline function. In Scenario 4 and 5 with more complex non-linear function \( C^Y(\cdot) \), GEAR-Bspline outperforms GEAR-linear in terms of smaller bias and higher empirical coverage probabilities of the 95% CI. For example, GEAR-Bspline achieves \( V(\hat{\beta}^G) = 2.43 \) in Scenario 4 (\( V(\beta_0) = 2.59 \)) with coverage probability 92.0% and \( V(\hat{\beta}^G) = 2.77 \) in Scenario 5 (\( V(\beta_0) = 3.03 \)) with coverage
Table 5.2: Empirical results under the GEAR for Scenario 3-5.

<table>
<thead>
<tr>
<th></th>
<th>GEAR-LINEAR</th>
<th>GEAR-BSPLINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_E = )</td>
<td>200 400 800</td>
<td>200 400 800</td>
</tr>
<tr>
<td>S3</td>
<td>( V(\beta_0) = 1.20 )</td>
<td>( V(\beta_G) )</td>
</tr>
<tr>
<td></td>
<td>( SE{\tilde{V}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( E{\tilde{\sigma}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( V(\tilde{\beta}_G) )</td>
<td>1.18 1.19 1.19</td>
</tr>
<tr>
<td>CP (%)</td>
<td>95.2 96.0 92.6</td>
<td>94.0 95.4 94.4</td>
</tr>
<tr>
<td>S4</td>
<td>( V(\beta_0) = 2.59 )</td>
<td>( V(\beta_G) )</td>
</tr>
<tr>
<td></td>
<td>( SE{\tilde{V}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( E{\tilde{\sigma}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( V(\tilde{\beta}_G) )</td>
<td>2.32 2.32 2.33</td>
</tr>
<tr>
<td>CP (%)</td>
<td>77.6 66.2 55.2</td>
<td>94.6 92.0 90.0</td>
</tr>
<tr>
<td>S5</td>
<td>( V(\beta_0) = 3.03 )</td>
<td>( V(\beta_G) )</td>
</tr>
<tr>
<td></td>
<td>( SE{\tilde{V}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( E{\tilde{\sigma}_{AIP}} )</td>
<td>( 0.02 ) 0.01 0.01</td>
</tr>
<tr>
<td></td>
<td>( V(\tilde{\beta}_G) )</td>
<td>2.30 2.32 2.32</td>
</tr>
<tr>
<td>CP (%)</td>
<td>31.6 17.4 11.8</td>
<td>96.0 92.4 87.8</td>
</tr>
</tbody>
</table>

probability 92.4% when \( N_E = N_U \), while GEAR-linear can hardly maintain an empirical coverage probability over one third in Scenario 5 due to the severe model misspecification. Note that because of the interaction between \( X \) and \( M \) in \( C^V(\cdot) \), the model assumption is still mildly violated even applying the GEAR-Bspline method. Thus, the empirical coverage probabilities of the 95% CI decreases as the sample size \( N_E \) increases.
5.5 Real Data Analysis

In this section, we illustrate our proposed method by application to the AIDS Clinical Trials Group Protocol 175 (ACTG 175) data. There are 1046 HIV-infected subjects enrolled in ACTG 175, who were randomized to two competitive antiretroviral regimens in equal proportions (Hammer et al. 1996): zidovudine (ZDV) + zalcitabine (ddC), and ZDV+didanosine (dDI). Denote ‘ZDV+ddC’ as treatment 0, versus ‘ZDV+dDI’ as treatment 1. The long-term outcome of interest (Y) is the mean CD4 count (cells/mm3) at 96 ± 5 weeks. A higher CD4 count usually indicates a stronger immune system. However, about one-third of the patients who received treatment 0 or 1 have missing long-term outcome, which form the experimental sample of interest. The rest dataset is used as the auxiliary sample.

In the experimental sample (NE = 376), 187 patients were randomized to treatment 0 and 189 patients to treatment 1, with the propensity score as constant π(x) ≡ 0.503. The auxiliary sample consists of NU = 670 subjects with observed long-term outcome. We consider r = 12 baseline covariates used in (Tsiatis et al. 2008): 1) four continuous variables: age (years), weight (kg), CD4 count (cells/mm3) at baseline, and CD8 count (cells/mm3) at baseline; 2) eight categorical variables: hemophilia, homosexual activity, history of intravenous drug use, Karnofsky score (scale of 0-100), race (0=white, 1=non-white), gender (0=female), antiretroviral history (0=naive, 1=experienced), and symptomatic status (0=asymptomatic). Intermediate outcomes contain CD4 count at 20 ± 5 weeks and CD8 count at 20 ± 5 weeks. It can be shown in the auxiliary data that intermediate outcomes are highly related to the long-term outcome via a linear regression of YU on {XU, MU}.

We apply our proposed ‘GEAR-linear’ and ‘GEAR-B-spline’ described in Section 5.4.2 to the ACTG 175 data, respectively. Here, to avoid the curse of high dimensionality, we only take the polynomial basis on the continuous variables with degree as 2. Reported in Table 5.3 are the estimated mean outcome for each treatment as \( \bar{Y}_{AIP}(0) \) and \( \bar{Y}_{AIP}(1) \), the estimated value \( \hat{\beta}_G \) with its estimated standard deviation \( \hat{\sigma}_{AIP} \), the 95% CI for the estimated value, and the number of assignments for each treatment.

It is clear that the proposed GEAR estimation procedure with the B-spline performs reasonably better than the linear procedure. Next, we focus on the results obtained from the GEAR-B-spline method in the experimental sample of interest. Our proposed GEAR-B-spline method achieves a value of 346.4 with a smaller standard deviation as 9.6 comparing to GEAR-linear (10.2) in the experimental sample. The GEAR with B-spline assigns 192 patients to ‘ZDV+dDI’ and 184 patients to ‘ZDV+ddC’, which is consistent with the competitive nature of these two treatments.
Table 5.3: Comparison results for ACTG 175 data.

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>B-spline</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{V}_{API}(0)$</td>
<td>327.8</td>
<td>325.5</td>
</tr>
<tr>
<td>$\hat{V}_{API}(1)$</td>
<td>334.0</td>
<td>329.0</td>
</tr>
<tr>
<td>$\hat{V}_{API}(\beta^G)$ [SD]</td>
<td>351.4 [10.2]</td>
<td>346.4 [9.6]</td>
</tr>
<tr>
<td>95% CI for $\hat{V}_{API}(\beta^G)$</td>
<td>(331.4, 371.3)</td>
<td>(327.7, 365.1)</td>
</tr>
<tr>
<td>ASSIGN TO ‘ZDV+DDC’</td>
<td>185</td>
<td>184</td>
</tr>
<tr>
<td>ASSIGN TO ‘ZDV+DDI’</td>
<td>191</td>
<td>192</td>
</tr>
</tbody>
</table>

5.6 Discussions

In this chapter, we proposed a new personalized optimal decision policy when the long-term outcome of interest cannot be observed. Theoretically, we gave the cubic convergence rate of our proposed GEAR, and derived the consistency and asymptotical distributions of the value function under the GEAR. Empirically, we validated our method, and examined the sensitivity of our proposed GEAR when the model is misspecified or when assumptions are violated.

There are several other possible extensions we may consider in future work. First, we only consider two treatment options in this work, while in applications it is common to have more than two options for decision making. Thus, a more general method with multiple treatments or even continuous decision marking is desirable. Second, we can extend our work to dynamic decision making, where the ultimate outcome of interest cannot be observed in the experimental sample but can be found in some auxiliary dataset.
REFERENCES


A.1 More on Tuning in L-JIL

For L-JIL, we choose \( \gamma_n \) and \( \lambda_n \) simultaneously via cross-validation. As we will show below, the use of cross-validation will not increase the computation complexity substantially in L-JIL.

To elaborate, let us revisit the proposed jump interval-learning in Algorithm 1. The most time consuming part lies in computing the ridge-type estimator

\[
\hat{\theta}_\mathcal{I}(\lambda_n) = \left( \sum_{i \in G_{k}} \overline{X}_i \overline{X}_i^\top \mathbb{I}(A_i \in \mathcal{I}) + n \lambda_n |\mathcal{I}| \mathbb{I}_p \right)^{-1} \left( \sum_{i \in G_{k}} \overline{X}_i Y_i \mathbb{I}(A_i \in \mathcal{I}) \right), \tag{A.1}
\]

where \( \mathbb{I}_p \) is the identity matrix with dimension \( p + 1 \), and the cost function

\[
\text{cost}(\mathcal{I}, \lambda_n) = \frac{1}{n} \sum_{i \in G_k} \mathbb{I}(A_i \in \mathcal{I}) \left\{ Y_i - \overline{X}_i^\top \hat{\theta}_\mathcal{I}(\lambda_n) \right\}^2,
\]

for any \( \mathcal{I} \in \{[l/m, r/m) : 1 \leq l < r < m \} \cup \{[l/m, 1] : 1 \leq l < m \} \).

To compute \( \{ \hat{\theta}_{\mathcal{I}, \mathcal{J}, \lambda_n, k} : \gamma_n \in \Gamma_n, \lambda_n \in \Lambda_n \} \), we need to calculate \( \{ \hat{\theta}_\mathcal{I}(\lambda_n) : \lambda_n \in \Lambda_n \} \) and
\{\text{cost}(\mathcal{J}, \lambda_n): \lambda_n \in \Lambda_n\} \text{ for any } \mathcal{J}. \text{ We first factorize the matrix } \sum_{i \in G_{-k}} \overline{X}_i \overline{X}_i^\top \mathbb{1}(A_i \in \mathcal{J}) \text{ as }

\sum_{i \in G_{-k}} \overline{X}_i \overline{X}_i^\top \mathbb{1}(A_i \in \mathcal{J}) = U \mathcal{T} U^\top,

according to the eigendecomposition, where } U \text{ is some } (p+1) \times (p+1) \text{ orthogonal matrix and } \mathcal{T} = \text{diag}(\tau_0, \tau_1, \ldots, \tau_p) \text{ is some diagonal matrix. Let } \phi = U^\top \{\sum_{i \in G_{-k}} \overline{X}_i \mathbb{1}(A_i \in \mathcal{J})\}. \text{ Then the set of estimators } \{\hat{\theta}_\mathcal{J}(\lambda_n): \lambda_n \in \Lambda_n\} \text{ can be calculated by }

\hat{\theta}_\mathcal{J}(\lambda_n) = U \text{diag}\{ (\tau_0 + n\lambda_n|\mathcal{J}|)^{-1}, (\tau_1 + n\lambda_n|\mathcal{J}|)^{-1}, \ldots, (\tau_p + n\lambda_n|\mathcal{J}|)^{-1}\} \phi,

simultaneously for all } \lambda_n.

Compared to separately inverting the matrix \sum_{i \in G_{-k}} \overline{X}_i \overline{X}_i^\top \mathbb{1}(A_i \in \mathcal{J}) + n\lambda_n|\mathcal{J}| \mathbb{E}_{p+1} \text{ in (A.1) for each } \lambda_n \text{ to compute } \{\hat{\theta}_\mathcal{J}(\lambda_n): \lambda_n \in \Lambda_n\}, \text{ the proposed method saves a lot of time especially for large } p. \text{ Similarly, based on the eigendecomposition, we have }

\[ n \text{cost } (\mathcal{J}, \lambda_n) = \sum_{i \in G_{-k}} Y_i^2 \mathbb{1}(A_i \in \mathcal{J}) \]

\[-2\phi^\top \text{diag}\{ (\tau_0 + n\lambda_n|\mathcal{J}|)^{-1}, (\tau_1 + n\lambda_n|\mathcal{J}|)^{-1}, \ldots, (\tau_p + n\lambda_n|\mathcal{J}|)^{-1}\} \phi \]

\[ + \phi^\top \text{diag}\{ \tau_0(\tau_0 + n\lambda_n|\mathcal{J}|)^{-2}, \tau_1(\tau_1 + n\lambda_n|\mathcal{J}|)^{-2}, \ldots, \tau_p(\tau_p + n\lambda_n|\mathcal{J}|)^{-2}\} \phi, \]

for all } \lambda_n \in \Lambda_n. \text{ This facilitates the computation of } \{\text{cost}(\mathcal{J}, \lambda_n): \lambda_n \in \Lambda_n\}.

After obtaining these cost functions, we can recursively compute the Bellman function } B(r, \lambda_n, \gamma_n) \text{ by }

\[ B(r, \lambda_n, \gamma_n) = \min_{j \in \mathcal{J}_r} \{ B(j, \lambda_n, \gamma_n) + \gamma_n + \text{cost}(\lfloor j/m, r/m, \lambda_n \rfloor) \}, \]

for all } r \geq 1, \lambda_n \in \Lambda_n \text{ and } \gamma_n \in \Gamma_n. \text{ Given the Bellman function, the set of estimators } \{\hat{\theta}_{\mathcal{J}, \gamma_n, \lambda_n, k}: \gamma_n \in \Gamma_n, \lambda_n \in \Lambda_n\} \text{ thus can be computed efficiently.}

### A.2 Technical Proofs

In the proofs, we use } c, C > 0 \text{ to denote some universal constants whose values are allowed to change from place to place. For any vector } \phi \in \mathbb{R}^q, \text{ we use } \phi^{(j)} \text{ to denote the } j\text{-th element of } \phi, \text{ for any } j \in \{1, \ldots, q\}. \text{ For any two positive sequences } \{a_n\}, \{b_n\}, a_n \asymp b_n \text{ means that } a_n \leq c b_n \text{ for some universal constant } c > 0
A.2.1 Proof of Theorem 1

We provide the proof for Theorem 1 in this section. We present an outline of the proof first. Let \( \delta_{\text{min}} = \min_{I \in \mathcal{P}} |I|/3 > 0 \). We divide the proof into four parts. In Part 1, we show that the following event occurs with probability at least \( 1 - O(n^{-2}) \),

\[
\max_{\tau \in \mathcal{J}(\mathcal{P}_0)} \min_{\hat{\tau} \in \mathcal{J}(c \mathcal{P})} |\hat{\tau} - \tau| < \delta_{\text{min}}. \tag{A.3}
\]

By the definition of \( \delta_{\text{min}} \), this implies that

\[
\Pr(|c \mathcal{P}| \geq |\mathcal{P}_0|) \geq 1 - O(n^{-2}). \tag{A.4}
\]

In Part 2, we show that

\[
\max_{\tau \in \mathcal{J}(\mathcal{P}_0)} \min_{\hat{\tau} \in \mathcal{J}(c \mathcal{P})} |\hat{\tau} - \tau| = O(n^{-1} \log n), \tag{A.5}
\]

with probability at least \( 1 - O(n^{-2}) \). This proves (ii) in Theorem 1. In Part 3, we prove

\[
\Pr(|\mathcal{P}| \leq |\mathcal{P}_0|) \geq 1 - O(n^{-2}). \tag{A.6}
\]

This together with (A.4) proves (i) in Theorem 1. In the last part, we show (iii) holds.

In the following, we first introduce some notations and auxiliary lemmas. Then, we present the proofs for Part 1, 2, 3 and 4.

Notations and technical lemmas: For any interval \( \mathcal{I} \subseteq [0, 1] \), define

\[
\hat{\theta}_\mathcal{I} = \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(A_i \in \mathcal{I}) \overline{X}_i \overline{X}_i^\top + \lambda_n |\mathcal{I}| \mathbb{E}_{\mathcal{P}_0} \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(A_i \in \mathcal{I}) \overline{X}_i Y_i \right),
\]

\[
\theta_{0, \mathcal{I}} = \left( \mathbb{E}(A \in \mathcal{I}) \overline{X} \overline{X}^\top \right)^{-1} \{ \mathbb{E}(A \in \mathcal{I}) \overline{X} Y \},
\]

where \( \overline{X} = (1, X^\top)^\top \). It is immediate to see that the definition of \( \hat{\theta}_\mathcal{I} \) here is consistent with the one defined in (2.3) for any \( \mathcal{I} \in \mathcal{P} \). In addition, under the model assumption in (2.13), the definition of \( \theta_{0, \mathcal{I}} \) here is consistent with the one defined in step function model \( \theta_0(a) = \sum_{\mathcal{I} \in \mathcal{P}_0} \theta_{0, \mathcal{I}} \mathbb{1}(a \in \mathcal{I}) \) for any \( a \in \mathcal{P}_0 \).
Let $\mathcal{I}(m)$ denote the set of intervals

\[ \mathcal{I}(m) = \{ [i_1/m, i_2/m) : \text{for some integers } i_1 \text{ and } i_2 \text{ that satisfy } 0 \leq i_1 < i_2 < m \} \]
\[ \cup \ \{ [i_3/m, 1) : \text{for some integers } i_3 \text{ that satisfy } 0 \leq i_3 < m \}. \]

Let $\{\tau_{0,k}\}_{k=1}^{K-1}$ with $0 < \tau_{0,1} < \tau_{0,2} < \cdots < \tau_{0,K-1} < 1$ be the locations of the true change points of $\theta_0(\cdot)$. Set $\tau_{0,0} = 0$, $\tau_{0,K} = 1$. We introducing the following lemmas.

**Lemma 1** Assume conditions in Theorem 1 are satisfied. Then there exist some constants $\bar{c}_0 > 0$, $c_0 \geq 1$ such that the following events occur with probability at least $1 - O(n^{-2})$: for any interval $\mathcal{J} \in \mathcal{I}(m)$ that satisfies $|\mathcal{J}| \geq \bar{c}_0 n^{-1} \log n$, we have

\begin{equation}
\| \hat{\theta}_\mathcal{J} - \theta_{0,\mathcal{J}} \|_2 \leq \frac{c_0 \sqrt{\log n}}{\sqrt{|\mathcal{J}|n}}, \tag{A.7}
\end{equation}
\begin{equation}
\left\| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) (Y_i - \bar{X}_i^T \theta_{0,\mathcal{J}}) \bar{X}_i \right\|_2 \leq \frac{c_0 \sqrt{|\mathcal{J}| \log n}}{\sqrt{n}}, \tag{A.8}
\end{equation}
\begin{equation}
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) (Y_i - \bar{X}_i^T \theta_{0}(A_i)) \bar{X}_i^T (\theta_{0}(A_i) - \theta_{0,\mathcal{J}}) \right| \leq \frac{c_0 \sqrt{|\mathcal{J}| \log n}}{\sqrt{n}}, \tag{A.9}
\end{equation}
\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \| \bar{X}_i^T (\theta_{0}(A_i) - \theta_{0,\mathcal{J}}) \|^2 \geq \frac{1}{c_0} \int_{\mathcal{J}} \| \theta_{0}(a) - \theta_{0,\mathcal{J}} \|^2 \, da - \frac{c_0 \sqrt{|\mathcal{J}| \log n}}{\sqrt{n}}, \tag{A.10}
\end{equation}
\begin{equation}
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) (|Y_i|^2 + \|\bar{X}_i\|^2_2) \leq c_0 \left( \frac{\sqrt{|\mathcal{J}| \log n}}{\sqrt{n}} + |\mathcal{J}| \right). \tag{A.11}
\end{equation}

In addition, for any $\mathcal{J} \in [0, 1]$, we have

\[ \| \theta_{0,\mathcal{J}} \|_2 \leq c_0. \tag{A.12} \]

**Lemma 2** Assume conditions in Theorem 1 are satisfied. Then there exist some constants $\bar{c}_1 > 0$, $c_1 \geq 1$ such that the following events occur with probability at least $1 - O(n^{-2})$: for any
interval $\mathcal{J} \in \mathcal{J}(m)$ that satisfies $\int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da \geq \tilde{c}_1 n^{-1} \log n,$

$$\left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{Y_i - \frac{X_i^T \theta_0(A_i)}{X_i^T \theta_0(A_i)}\} X_i^T \{\theta_0(A_i) - \theta_{0,\mathcal{J}}\} \right| \leq c_1 \sqrt{n \int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da \log n},$$

(A.13)

$$\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \|X_i^T \{\theta_0(A_i) - \theta_{0,\mathcal{J}}\}\|^2 \geq \frac{n}{c_1} \int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da - c_1 \sqrt{n \int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da \log n}.$$  (A.14)

**Lemma 3** Assume conditions in Theorem 1 are satisfied. Then for sufficiently large $n$ and any interval $\mathcal{J} \subseteq [0, 1]$ of the form $[i_1, i_2)$ or $[i_1, i_2]$ with $i_2 = 1$ that satisfies $\int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da = c_n$ for some sequence $\{c_n\}$ such that $c_n \geq 0, \forall n$ and $c_n \rightarrow 0$ as $n \rightarrow \infty$, we have either $\tau_{0,k-1} \leq i_1 \leq i_2 \leq \tau_{0,k}$ for some integer $k$ such that $1 \leq k \leq K$ or

$$\tau_{0,k-2} \leq i_1 < \tau_{0,k-1} < i_2 \leq \tau_{0,k} \text{ and } \min_{j \in \{1, 2\}} |i_j - \tau_{0,k-1}| \leq c_2 c_n,$$

for some integer $k$ such that $2 \leq k \leq K$ and some constant $c_2 > 0$, or

$$\tau_{0,k-3} \leq i_1 < \tau_{0,k-2} < \tau_{0,k-1} < i_2 \leq \tau_{0,k} \text{ and } \max_{j \in \{1, 2\}} |i_j - \tau_{0,k-3+j}| \leq c_2 c_n,$$

for some integer $k$ such that $3 \leq k \leq K$ and some constant $c_2 > 0$.

In addition, the following events occur with probability at least $1 - O(n^{-2})$: for any interval $\mathcal{J} \in \mathcal{J}(m)$ that satisfies $\int_{\mathcal{J}} \|\theta_0(a) - \theta_{0,\mathcal{J}}\|^2_2 da \leq \tilde{c}_1 n^{-1} \log n$, we have

$$\left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{Y_i - X_i^T \theta_0(A_i)\} X_i^T \{\theta_0(A_i) - \theta_{0,\mathcal{J}}\} \right| \leq \tilde{c}_2 \log n,$$

(A.15)

for some constant $\tilde{c}_2 > 0$.

**Lemma 4** Under the conditions in Theorem 1, the following events occur with probability at least $1 - O(n^{-2})$: there exists some constant $\tilde{c}_3 > 0$ such that $\min_{\mathcal{J} \in \mathcal{J}} |\mathcal{J}| \geq \tilde{c}_3 \gamma_n$.

**Part 1:** Assume $|\mathcal{P}_0| > 1$. Otherwise, (A.3) trivially hold. Consider the partition $\mathcal{P} = \{[0, 1]\}$
which consists of a single interval and a zero vector $0_{p+1}$. By definition, we have
\[
\sum_{S \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \hat{\theta}_S)^2 + n \lambda_n |S| \|\hat{\theta}_S\|_2^2 \right) + n \gamma_n |S| \leq \sum_{i=1}^{n} (Y_i - \overline{X}_i 0_{p+1})^2 + n \lambda_n \|0_{p+1}\|_2^2 + n \gamma_n = \sum_{i=1}^{n} Y_i^2 + n \gamma_n.
\]

In view of (A.11), we obtain with probability at least $1 - O(n^{-2})$,
\[
\sum_{S \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \hat{\theta}_S)^2 + n \lambda_n |S| \|\hat{\theta}_S\|_2^2 \right) + n \gamma_n (|\mathcal{S}| - 1) \leq c_0 n \left( \frac{\sqrt{\log n}}{\sqrt{n}} + 1 \right).
\]

This implies that under the event defined in (A.11), we have for sufficiently large $n$,
\[
\gamma_n (|\mathcal{S}| - 1) \leq c_0 \left( \frac{\sqrt{\log n}}{\sqrt{n}} + 1 \right),
\]
and hence
\[
|\mathcal{S}| \leq 2 c_0 \gamma_n^{-1}, \tag{A.16}
\]
for sufficiently large $n$.

Under the event defined in Lemma 4, we have $\min_{S \in \mathcal{P}} |S| \geq c_0 n^{-1} \log n$ for sufficiently large $n$, since $\gamma_n \gg n^{-1} \log n$. Thus, with probability at least $1 - O(n^{-2})$, the events defined in (A.7)-(A.11) hold for any interval $S \in \mathcal{P}$.

Notice that
\[
\sum_{S \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \hat{\theta}_S)^2 + n \lambda_n |S| \|\hat{\theta}_S\|_2^2 \right) + n \gamma_n |S| \geq \sum_{S \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \theta_{0,S})^2 \tag{A.17}
\]
\[
\geq \sum_{S \in \mathcal{P}} \left[ \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \hat{\theta}_S)^2 + n \gamma_n |S| \right] + \sum_{S \in \mathcal{P}} \left[ \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \theta_{0,S})^2 \right]
\]
\[
+ \sum_{S \in \mathcal{P}} \left[ \sum_{i=1}^{n} \mathbb{I}(A_i \in S)[ \overline{X}_i (\hat{\theta}_S - \theta_{0,S})]^2 \right] - 2 \sum_{S \in \mathcal{P}} \left[ \sum_{i=1}^{n} \mathbb{I}(A_i \in S)(Y_i - \overline{X}_i \theta_{0,S}) \overline{X}_i (\hat{\theta}_S - \theta_{0,S}) \right].
\]
By (A.7) and (A.8), we obtain that

\[ \eta_3 \leq 2 \sum_{I \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \bar{X}_i^T \theta_{0,I}) \bar{X}_i \right) \| \hat{\theta}_{\mathcal{I}} - \theta_{0,I} \|_2 \]

\[ \leq \sum_{I \in \mathcal{P}} 2c_0^2 \log n \leq 2c_0^2 \log n |\mathcal{P}|, \]

with probability at least \(1 - O(n^{-2})\).

Since \(\gamma_n \gg n^{-1} \log n\), \(\eta_2 \geq 0\), for sufficiently large \(n\), we have with probability at least \(1 - O(n^{-2})\),

\[ \sum_{A \in \mathcal{A}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (Y_i - \bar{X}_i^T \hat{\theta}_A)^2 + n \lambda_n |\mathcal{A}| \| \hat{\theta}_A \|_2^2 \right) + n \gamma_n |\mathcal{A}| \geq \eta_1. \]  

(A.19)

Notice that

\[ \eta_1 = \sum_{A \in \mathcal{A}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (Y_i - \bar{X}_i^T \theta_{0,A})^2 + \sum_{A \in \mathcal{A}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (\bar{X}_i^T \theta_{0,A} - \bar{X}_i^T \theta_{0,I})^2 \]

\[ = \sum_{A \in \mathcal{A}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (Y_i - \bar{X}_i^T \theta_{0,A}) + \sum_{A \in \mathcal{A}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (\bar{X}_i^T \theta_{0,A} - \bar{X}_i^T \theta_{0,I}) \]

\[ + 2 \sum_{A \in \mathcal{A}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{A}) (Y_i - \bar{X}_i^T \theta_{0,A}) (\bar{X}_i^T \theta_{0,A} - \bar{X}_i^T \theta_{0,I}). \]

Under the events defined in (A.9) and (A.10), it follows that

\[ \eta_1 \geq \eta_4 + n \sum_{A \in \mathcal{A}} \frac{1}{c_0} \int_{\mathcal{A}} \| \theta_0(a) - \theta_{0,I} \|_2^2 da - 2c_0 \sum_{A \in \mathcal{A}} |\mathcal{A}| \| \theta_0(a) - \theta_{0,I} \|_2 |\mathcal{A}| \]

\[ \geq \eta_4 + n \sum_{A \in \mathcal{A}} \frac{1}{c_0} \int_{\mathcal{A}} \| \theta_0(a) - \theta_{0,I} \|_2^2 da - 2c_0 \sqrt{|\mathcal{A}| n \log n}, \]

where the last inequality is due to Cauchy-Schwarz inequality. By (A.16) and the condition that \(\gamma_n \gg n^{-1} \log n\), we obtain

\[ \eta_1 \geq \eta_4 + n \sum_{A \in \mathcal{A}} \frac{1}{c_0} \int_{\mathcal{A}} \| \theta_0(a) - \theta_{0,I} \|_2^2 da + o(n), \]

(A.20)
with probability at least \(1 - O(n^{-2})\). Notice that

\[
\eta_4 = \sum_{i=1}^{n} \{Y_i - \overline{X}_i \theta_0(A_i)\}^2 = \sum_{\mathcal{I} \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \theta_0,\mathcal{I})^2.
\]

Combining (A.19) with (A.20), we've shown that

\[
\sum_{\mathcal{I} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \tilde{\theta})^2 + n \lambda_n |\mathcal{I}| \|\tilde{\theta}\|_2^2 \right) + n\gamma_n |\mathcal{P}|
\]

\[
\geq \sum_{\mathcal{I} \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \theta_0,\mathcal{I})^2 + n \sum_{\mathcal{I} \in \mathcal{P}} \frac{1}{c_0} \int_\mathcal{I} \|\theta_0(a) - \theta_0,\mathcal{I}\|_2^2 da + o(n),
\]

with probability at least \(1 - O(n^{-2})\). By (A.12) and the condition that \(\lambda_n = O(n^{-1} \log n)\), \(\gamma_n = o(1)\), this further implies

\[
\sum_{\mathcal{I} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \tilde{\theta})^2 + n \lambda_n |\mathcal{I}| \|\tilde{\theta}\|_2^2 \right) + n\gamma_n |\mathcal{P}|
\]

\[
\geq \sum_{\mathcal{I} \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \theta_0,\mathcal{I})^2 + n \lambda_n |\mathcal{I}| \|\theta_0,\mathcal{I}\|_2^2 + n\gamma_n |\mathcal{P}| + n \sum_{\mathcal{I} \in \mathcal{P}} \frac{1}{c_0} \int_\mathcal{I} \|\theta_0(a) - \theta_0,\mathcal{I}\|_2^2 da + o(n).
\]

(A.21)

For any integer \(k\) such that \(1 \leq k \leq K - 1\), let \(\tau_{0,k}^* = i/m\) for some integer \(i\) and that \(|\tau_{0,k} - \tau_{0,k}^*| < m^{-1}\). Denoted by \(\mathcal{P}^*\) the oracle partition formed by the change point locations \(\{\tau_{0,k}^*\}_{k=1}^{K-1}\). Set \(\tau_{0,0}^* = 0\), \(\tau_{0,K}^* = 1\) and \(\theta_{(\tau_{0,k-1}, \tau_{0,k})}^* = \theta_{0|\tau_{0,k-1}, \tau_{0,k}}\) for \(1 \leq k \leq K - 1\) and \(\theta_{(\tau_{0,k-1}, 1]}^* = \theta_{0|\tau_{0,K-1}, 1]}\). Let \(\Delta_k = [\tau_{0,k-1}, \tau_{0,k}) \cap [\tau_{0,k-1}, \tau_{0,k}]^c\) for \(1 \leq k \leq K - 1\) and \(\Delta_K = [\tau_{0,K-1}, 1]\). The length of each interval \(\Delta_k\) is at most \(m^{-1}\). Since \(m > n\), we have \(m^{-1} \ll \tilde{c}_0 n^{-1} \log n\). For any \(k\) and sufficiently large \(n\), we can find an interval \(\mathcal{I} \in \mathcal{I}(m)\) with length between \(\tilde{c}_0 n^{-1} \log n\) and \(2\tilde{c}_0 n^{-1} \log n\) that
covers $\Delta_k$. It follows that

$$
\left\{ \sum_{\mathscr{A} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathscr{A})(Y_i - \overline{X_i} \theta^*_{\mathscr{A}})^2 + n\lambda_n |\mathscr{A}| |\theta^*_{\mathscr{A}}|^{2} \right) + n\gamma_n |\mathcal{P}^*| \right\} 
- \left\{ \sum_{\mathscr{A} \in \mathcal{P}_0} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathscr{A})(Y_i - \overline{X_i} \theta_{0,\mathscr{A}})^2 + n\lambda_n |\mathscr{A}| |\theta_{0,\mathscr{A}}|^{2} \right) + n\gamma_n |\mathcal{P}_0| \right\}
\leq n\lambda_n \sup_{\mathscr{A} \subseteq \{0,1\}} |\theta_{0,\mathscr{A}}|^{2} + \sum_{k=1}^{K} \sum_{i=1}^{n} \mathbb{I}(A_i \in \Delta_k) \left( Y_i^2 + \sup_{\mathscr{A} \subseteq \{0,1\}} |\theta_{0,\mathscr{A}}|^{2} \right) \right|_{\mathscr{A}} \overline{X_i}^{2}
\leq n\lambda_n \sup_{\mathscr{A} \subseteq \{0,1\}} |\theta_{0,\mathscr{A}}|^{2} + K \sup_{\mathscr{A} \subseteq \{0,1\}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathscr{A}) \left( Y_i^2 + \sup_{\mathscr{A} \subseteq \{0,1\}} |\theta_{0,\mathscr{A}}|^{2} \right) \overline{X_i}^{2}
\leq c_0^2 n\lambda_n + K(c_0^2 + 1)c_0 (\sqrt{2c_0} + 2\bar{c}_0) \log n = O(\log n) = o(n).
$$

(A.23)

By definition, we have

$$
\sum_{\mathscr{A} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathscr{A})(Y_i - \overline{X_i} \tilde{\theta}_{\mathscr{A}})^2 + n\lambda_n |\mathscr{A}| |\tilde{\theta}_{\mathscr{A}}|^{2} \right) + n\gamma_n |\mathcal{P}|
\leq \sum_{\mathscr{A} \in \mathcal{P}^*} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathscr{A})(Y_i - \overline{X_i} \theta^*_{\mathscr{A}})^2 + n\lambda_n |\mathscr{A}| |\theta^*_{\mathscr{A}}|^{2} \right) + n\gamma_n |\mathcal{P}^*|.
$$

In view of (A.21) and (A.23), we obtain that

$$
\sum_{\mathscr{A} \in \mathcal{P}} \int \|\theta_0(a) - \theta_{0,\mathscr{A}}\|_{L}^2 d\alpha = o(1),
$$

(A.24)

with probability at least $1 - O(n^{-2})$. We now show (A.3) holds under the event defined in (A.24). Otherwise, there exists some $\tau_0 \in J(\mathcal{P}_0)$ such that $|\hat{\tau} - \tau_0| \geq \delta_{\min}$, for all $\hat{\tau} \in J(\mathcal{P}_0)$. 

119
Under the event defined in (A.24), we obtain that
\[
\int_{\tau_0 - \delta_{\min}}^{\tau_0 + \delta_{\min}} \| \theta_0(a) - \theta_0([\tau_0 - \delta_{\min}, \tau_0 + \delta_{\min}]) \|^2_2 \, da = o(1). \tag{A.25}
\]

On the other hand, since \( \theta_0(a) \) is a constant function on \([\tau_0 - \delta_{\min}, \tau_0)\) or \([\tau_0, \tau_0 + \delta_{\min})\), we have
\[
\int_{\tau_0 - \delta_{\min}}^{\tau_0 + \delta_{\min}} \| \theta_0(a) - \theta_0([\tau_0 - \delta_{\min}, \tau_0 + \delta_{\min}]) \|^2_2 \, da
\geq \min_{\theta \in \mathbb{R}^p \setminus \{0\}} \left( \delta_{\min} \| \theta_0([\tau_0 - \delta_{\min}, \tau_0]) - \theta \|^2 + \delta_{\min} \| \theta_0([\tau_0, \tau_0 + \delta_{\min}]) - \theta \|^2 \right)
\geq \frac{\delta_{\min}}{2} \| \theta_0([\tau_0 - \delta_{\min}, \tau_0]) - \theta_0([\tau_0, \tau_0 + \delta_{\min}]) \|^2_2 \geq \frac{\delta_{\min} \kappa_0^2}{2},
\]
where
\[
\kappa_0 \equiv \min_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}_0} \| \theta_0, \mathcal{S}_1 - \theta_0, \mathcal{S}_2 \|_2 > 0.
\]

This apparently violates (A.25). (A.3) thus holds with probability at least \(1 - O(n^{-2})\).

**Part 2:** By (A.17) and (A.18), we have with probability at least \(1 - O(n^{-2})\) that
\[
\sum_{\mathcal{S} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})(Y_i - \overline{X}_i^T \hat{\theta}_{\mathcal{S}})^2 + n \lambda_n |\mathcal{S}| \| \hat{\theta}_{\mathcal{S}} \|^2_2 \right) + n \gamma_n |\mathcal{P}| \geq \eta_1 + n \gamma_n |\mathcal{P}| - 2 c_0^2 |\mathcal{P}| \log n.
\]

Notice that
\[
\eta_1 = \eta_4 + 2 \sum_{\mathcal{S} \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})(Y_i - \overline{X}_i^T \theta_0(A_i)) [\overline{X}_i^T \theta_0(A_i) - \overline{X}_i^T \theta_{0, \mathcal{S}}]
+ \sum_{\mathcal{S} \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})(\overline{X}_i^T \theta_0(A_i) - \overline{X}_i^T \theta_{0, \mathcal{S}})^2.
\]

Denoted by \( \mathcal{I}(m) \) the set of intervals \( \mathcal{S} \in \mathcal{I}(m) \) with \( \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0, \mathcal{S}} \|^2_2 \, da \geq \tilde{c}_1 n^{-1} \log n. \)
Under the events defined in Lemma 2 and 3, we have

\[ \eta_1 \geq \eta_4 + 2 \sum_{\mathcal{J} \in \mathcal{D}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - X_i^\top \theta_0(A_i) \} \{ X_i^\top \theta_0(A_i) - X_i^\top \theta_0,\mathcal{J} \} + \sum_{\mathcal{J} \in \mathcal{D}, \mathcal{I} \in \mathcal{I}(m)} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ X_i^\top \theta_0(A_i) - X_i^\top \theta_0,\mathcal{J} \}^2 \geq \eta_4 - 2 \bar{c}_2 |\mathcal{D}| \log n \]

To summarize, we've shown that with probability at least \( 1 - O(n^{-2}) \),

\[ \sum_{\mathcal{J} \in \mathcal{D}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - X_i^\top \theta_0,\mathcal{J} \}^2 + n \lambda_n |\mathcal{J}| \| \hat{\theta}_0,\mathcal{J} \|_2^2 \right) + n \gamma_n |\mathcal{D}| \geq \sum_{\mathcal{J} \in \mathcal{D}, \mathcal{I} \in \mathcal{I}(m)} \left( \frac{n}{c_1} \int_{\mathcal{J}} \| \theta_0(a) - \theta_0,\mathcal{J} \|_2^2 da - 3c_1 n \int_{\mathcal{J}} \| \theta_0(a) - \theta_0,\mathcal{J} \|_2^2 da \log n \right) + \eta_4 + n \gamma_n |\mathcal{D}| - 2(\bar{c}_0^2 + \bar{c}_2) |\mathcal{D}| \log n. \quad (A.26) \]

It follows from (A.22) and (A.23) that

\[ \eta_4 + n \lambda_n \sup_{\mathcal{J} \in \mathcal{D}_0} \| \theta_0,\mathcal{J} \|_2^2 + n \gamma_n |\mathcal{D}_0| \geq \left\{ \sum_{\mathcal{J} \in \mathcal{D}_0} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - X_i^\top \theta_0^*,\mathcal{J} \}^2 + n \lambda_n |\mathcal{J}| \| \theta_0^*,\mathcal{J} \|_2^2 \right) + n \gamma_n |\mathcal{D}_0| \right\} - c_0^* \log n, \]

for some constants \( c_0^* > 0 \), with probability at least \( 1 - O(n^{-2}) \). By (A.12) and the condition that \( \lambda_n = O(n^{-1} \log n) \), there exists some constant \( c_1^* > c_0^* \) such that

\[ \eta_4 + n \gamma_n |\mathcal{D}_0| \geq \left\{ \sum_{\mathcal{J} \in \mathcal{D}_0} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - X_i^\top \theta_0^*,\mathcal{J} \}^2 + n \lambda_n |\mathcal{J}| \| \theta_0^*,\mathcal{J} \|_2^2 \right) + n \gamma_n |\mathcal{D}_0| \right\} - c_1^* \log n, \]

with probability at least \( 1 - O(n^{-2}) \). In view of (A.26), we've shown that with probability at
least $1 - O(n^{-2})$,

\[
\sum_{\mathcal{I} \in \mathcal{P}} \left( \sum_{i=1}^{n} \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \bar{X}_i^T \hat{\theta}_i^*)^2 + n\lambda_n |\mathcal{I}||\hat{\theta}_i^*||_{2}^{2} \right) + n\gamma_n |\mathcal{P}| \quad \text{(A.27)}
\]

\[ \geq \sum_{\mathcal{I} \in \mathcal{P}, \mathcal{I} \in \mathcal{T}(m)} \left( \frac{n}{c_1} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da - 3c_1 \sqrt{n} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da \log n \right) \]

\[ + \sum_{\mathcal{I} \in \mathcal{P}, \mathcal{I} \in \mathcal{T}(m)} \left( \frac{n}{c_1} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da - 3c_1 \sqrt{n} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da \log n \right) \]

\[ + \sum_{\mathcal{I} \in \mathcal{P}, \mathcal{I} \in \mathcal{T}(m)} \left( \frac{n}{c_1} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da - 3c_1 \sqrt{n} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da \log n \right) \]

By definition,

\[ \sum_{\mathcal{I} \in \mathcal{P}} \left( \sum_{i=1}^{n} \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \bar{X}_i^T \hat{\theta}_i^*)^2 + n\lambda_n |\mathcal{I}||\hat{\theta}_i^*||_{2}^{2} \right) + n\gamma_n |\mathcal{P}| \]

\[ \leq \sum_{\mathcal{I} \in \mathcal{P}} \left( \sum_{i=1}^{n} \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \bar{X}_i^T \hat{\theta}_i^*)^2 + n\lambda_n |\mathcal{I}||\hat{\theta}_i^*||_{2}^{2} \right) + n\gamma_n |\mathcal{P}^*|. \]

Thus, we have with probability at least $1 - O(n^{-2})$,

\[ \sum_{\mathcal{I} \in \mathcal{P}, \mathcal{I} \in \mathcal{T}(m)} \left( \frac{n}{c_1} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da - 3c_1 \sqrt{n} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da \log n \right) \]

\[ \leq (2c_0^2 + 2\tilde{c}_2)^2 |\mathcal{P}| \log n + c_1^* \log n + n\gamma_n |\mathcal{P}_0| - n\gamma_n |\mathcal{P}|, \]

and hence,

\[ \sum_{\mathcal{I} \in \mathcal{P}, \mathcal{I} \in \mathcal{T}(m)} \left( \frac{n}{c_1} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da - 3c_1 \sqrt{n} \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|_2^2 da \log n \right) \]

\[ \leq (2c_0^2 + 2\tilde{c}_2 + 9c_1/4) |\mathcal{P}| \log n + c_1^* \log n + n\gamma_n |\mathcal{P}_0| - n\gamma_n |\mathcal{P}|. \]
$2c_0^2 + 2\tilde c_2 + 9c_1/4$, $n|\mathcal{P}_0|\gamma_n \geq 2c_1^* \log n$ and hence

$$(2c_0^2 + 2\tilde c_2 + 9c_1/4)|\mathcal{F}| \log n + c_1^* \log n + n\gamma_n|\mathcal{P}_0| - n\gamma_n|\mathcal{F}|$$

\leq (2c_0^2 + 2\tilde c_2 + 9c_1/4)|\mathcal{F}| \log n + c_1^* \log n - n\gamma_n|\mathcal{P}_0|/2$$

\leq (2c_0^2 + 2\tilde c_2 + 9c_1/4)|\mathcal{F}| \log n - n\gamma_n|\mathcal{F}|/4 \leq 0,$$

When $|\mathcal{P}_0| \leq |\mathcal{F}| \leq 2|\mathcal{P}_0|$, we have

$$(2c_0^2 + 2\tilde c_2 + 9c_1/4)|\mathcal{F}| \log n + c_1^* \log n + n\gamma_n|\mathcal{P}_0| - n\gamma_n|\mathcal{F}|$$

\leq 2(2c_0^2 + 2\tilde c_2 + 9c_1/4)|\mathcal{P}_0| \log n + c_1^* \log n.$$

In view of (A.28), we have with probability at least $1 - O(n^{-2})$,

$$\sum_{\mathcal{I} \in \mathcal{F}, \mathcal{I} \in \Xi(m)} \frac{n}{c_1} \left( \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|^2_2 da - \frac{3c_1}{2} n^{-1/2} \sqrt{\log n} \right)^2 \leq c \log n,$$

for some constant $c > 0$. Thus, with probability at least $1 - O(n^{-2})$, we have

$$\int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|^2_2 da = O(n^{-1} \log n),$$

for any $\mathcal{I} \in \mathcal{F} \cap \Xi(m)$. By the definition of $\Xi(m)$, we obtain that with probability at least $1 - O(n^{-2})$,

$$\int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,\mathcal{I}}\|^2_2 da = O(n^{-1} \log n), \ \forall \mathcal{I} \in \mathcal{F}. \ \ \ \ \ (A.29)$$

Consider a given change point $\tau \in \mathcal{P}_0$, there exists an interval $\mathcal{I} \in \mathcal{F}$ of the form $[i_1, i_2)$ or $[i_1, i_2]$ with $i_2 = 1$ such that $i_1 \leq \tau < i_2$. Under the event defined in (A.29), it follows from Lemma 3 such that $\min(|i_1 - \tau|, |i_2 - \tau|) = O(n^{-1} \log n)$. This proves (A.5).

**Part 3:** Using similar arguments in proving (A.15), we can show that the following events occur with probability at least $1 - O(n^{-2})$: for any interval $\mathcal{I} \in \mathcal{F}$, we have

$$\left| \sum_{i=1}^n \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - \overline{X}_i^T \theta_0(A_i) \} \overline{X}_i^T \{ \theta_0(A_i) - \theta_{0,\mathcal{I}} \} \right| \leq C \log n,$$

for some constant $C > 0$. 

123
By (A.29), using similar arguments in proving (A.26) and (A.27), we can show the following event occurs with probability at least \(1 - O(n^{-2})\),

\[
\sum_{\mathcal{J} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J})(Y_i - \mathbf{X}_i^\top \hat{\theta}_\mathcal{J})^2 + n \lambda_n |\mathcal{J}| \|\hat{\theta}_\mathcal{J}\|_2^2 \right) + n \gamma_n |\mathcal{J}^*| \]
\[
\geq \left\{ \sum_{\mathcal{J} \in \mathcal{P}^*} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J})(Y_i - \mathbf{X}_i^\top \theta_\mathcal{J}^*)^2 + n \lambda_n |\mathcal{J}| \|\theta_\mathcal{J}^*\|_2^2 \right) + n \gamma_n |\mathcal{J}^*| \right\}
\]
\[
+ n \gamma_n |\mathcal{P}| - C |\mathcal{J}^*| \log n - n \gamma_n |\mathcal{P}_0|,
\]

for some constant \(C > 0\). By definition,

\[
\sum_{\mathcal{J} \in \mathcal{P}} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J})(Y_i - \mathbf{X}_i^\top \hat{\theta}_\mathcal{J})^2 + n \lambda_n |\mathcal{J}| \|\hat{\theta}_\mathcal{J}\|_2^2 \right) + n \gamma_n |\mathcal{J}|
\]
\[
\leq \sum_{\mathcal{J} \in \mathcal{P}^*} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J})(Y_i - \mathbf{X}_i^\top \theta_\mathcal{J}^*)^2 + n \lambda_n |\mathcal{J}| \|\theta_\mathcal{J}^*\|_2^2 \right) + n \gamma_n |\mathcal{J}^*|.
\]

Thus, we have with probability at least \(1 - O(n^{-2})\),

\[
n \gamma_n |\mathcal{P}| - C |\mathcal{J}^*| \log n - n \gamma_n |\mathcal{P}_0| \leq 0.
\]

Since \(\gamma_n \gg n^{-1} \log n\), the above event occurs only when \(|\mathcal{J}^*| \leq |\mathcal{P}_0|\). To see this, notice that if \(|\mathcal{J}^*| > |\mathcal{P}_0|\), we have

\[
n \gamma_n - C \log n - \frac{C \log n}{|\mathcal{P}|} - n \gamma_n |\mathcal{J}_0| \geq n \gamma_n - C \log n - \frac{C \log n}{|\mathcal{P}_0| + 1} - n \gamma_n |\mathcal{J}_0|
\]
\[
= \frac{n \gamma_n}{|\mathcal{J}_0| + 1} - C \log n - \frac{C \log n}{|\mathcal{P}_0| + 1} \gg 0,
\]

since \(\gamma_n \gg n^{-1} \log n\). This proves (A.6).

**Part 4:** In the first three parts, we’ve shown that

\[
|\mathcal{J}_0| = |\mathcal{P}_0| \quad \text{and} \quad \max \min_{\tau \in |\mathcal{J}_0|} |\mathcal{J}_0| \leq |\mathcal{J}| = O(n^{-1} \log n), \quad (A.30)
\]

with probability tending to 1. For sufficiently large \(n\), the event defined in (A.30) implies that \(|\mathcal{J}| \geq \bar{c} n^{-1} \log n\) for any \(\mathcal{J} \in \mathcal{P}\). Thus, it follows from Lemma 1 that the following
occurs with probability at least $1 - O(n^{-2})$: for any $\mathcal{I} \in \tilde{\mathcal{P}}$, we have

$$|\mathcal{I}| \|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2^2 \leq c_0^2 n^{-1} \log n. \quad \text{(A.31)}$$

Under the events defined in (A.29), (A.30) and (A.31), we have

$$\int_0^1 \|	ilde{\theta}(a) - \theta(a)\|_2^2 da = \sum_{\mathcal{I} \in \mathcal{P}} \int \|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2^2 da = \sum_{\mathcal{I} \in \mathcal{P}} \int \|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}} + \theta_{0, \mathcal{I}} - \theta(a)\|_2^2 da$$

$$= \sum_{\mathcal{I} \in \mathcal{P}} \int \|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2^2 da + \sum_{\mathcal{I} \in \mathcal{P}} \int \|	heta_{0, \mathcal{I}} - \theta(a)\|_2^2 da + 2 \sum_{\mathcal{I} \in \mathcal{P}} \int (\tilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}) \{\theta_{0, \mathcal{I}} - \theta(a)\} da$$

$$\leq 2 \sum_{\mathcal{I} \in \mathcal{P}} \int \|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2^2 da + 2 \sum_{\mathcal{I} \in \mathcal{P}} \int \|	heta_{0, \mathcal{I}} - \theta(a)\|_2^2 da = O(|\mathcal{P}| n^{-1} \log n) = O(|\mathcal{P}| n^{-1} \log n),$$

where the first inequality is due to Cauchy-Schwarz inequality. This proves (iii). The proof is hence completed.

### A.2.2 Proof of Lemma 1

**Proof of (A.7):** By definition, we have

$$\|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2$$

$$\leq \left\| \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i \bar{X}_i^T + \lambda_n \mathcal{I} | \mathcal{I} \mathcal{E}_{p+1} \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i Y_i - \mathbb{E}(A \in \mathcal{I}) \bar{X} Y \right) \right\|_2$$

$$+ \left\| \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i \bar{X}_i^T + \lambda_n \mathcal{I} | \mathcal{I} \mathcal{E}_{p+1} \right)^{-1} - \left( \mathbb{E}(A \in \mathcal{I}) \bar{X} \bar{X}^T \right)^{-1} \right\|_2 \left\{ \mathbb{E}(A \in \mathcal{I}) \bar{X} Y \right\}$$

$$\leq \left\| \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i \bar{X}_i^T + \lambda_n \mathcal{I} | \mathcal{I} \mathcal{E}_{p+1} \right)^{-1} \right\|_2 \left\| \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i Y_i - \mathbb{E}(A \in \mathcal{I}) \bar{X} Y \right) \right\|_2$$

$$+ \left\| \left( \frac{1}{n} \sum_{i=1}^n I(A_i \in \mathcal{I}) \bar{X}_i \bar{X}_i^T + \lambda_n \mathcal{I} | \mathcal{I} \mathcal{E}_{p+1} \right)^{-1} - \left( \mathbb{E}(A \in \mathcal{I}) \bar{X} \bar{X}^T \right)^{-1} \right\|_2 \left\| \mathbb{E}(A \in \mathcal{I}) \bar{X} Y \right\|_2.$$

It follows from Cauchy-Schwarz inequality that

$$\|	ilde{\theta}_{\mathcal{I}} - \theta_{0, \mathcal{I}}\|_2^2 \leq 2 \eta_1^2(I) \eta_2^2(I) + 2 \eta_3^2(I) \eta_4^2(I). \quad \text{(A.32)}$$
In the following, we provide upper bounds for

$$\max_{\mathcal{I} \in \mathcal{I}(n)} \eta_j(\mathcal{I}),$$

for \( j = 1, 2, 3, 4 \), where the constant \( \tilde{c}_0 \) will be specified later. The uniform convergence rates of \( \| \hat{\theta}_\mathcal{I} - \theta_{0,\mathcal{I}} \|_2 \) can thus be derived.

Without loss of generality, assume the constant \( \omega \) in Condition (A4) is greater than or equal to \( \log^{-1/2} 2 \). Then, we have \( \exp(1/\omega^2) \leq \exp(\log 2) = 2 \) and hence \( \| 1 \|_{q,|A|} \leq \omega \). Therefore, we have \( \max_{j \in \{1, \ldots, p+1\}} \| \bar{X}^{(j)} \|_{q,|A|} \leq \omega \), almost surely. By the definition of the conditional Orlicz norm, this implies that

$$E \left\{ 1 + \sum_{q=1}^{+\infty} \frac{|\bar{X}^{(j)}|^2 q}{\omega^{2q} q!} \right\} \leq 2, \quad \forall j \in \{1, \ldots, p+1\},$$

almost surely, and hence

$$E(|\bar{X}^{(j)}|^{2q}|A) \leq q! \omega^{2q}, \quad \forall j \in \{1, \ldots, p+1\}, q = 1, 2, \ldots \quad (A.33)$$

By Cauchy-Schwarz inequality, we obtain that

$$E(|\bar{X}^{(j)}|^{2q}|A) \leq \sqrt{E(|\bar{X}^{(j)}|^{2q}|A)E(|\bar{X}^{(j)}|^{2q}|A)} \leq q! \omega^{2q}, \quad (A.34)$$

for any \( j_1, j_2 \in \{1, \ldots, p+1\} \) and any integer \( q \geq 1 \), almost surely.

Since \( A \) has a bounded probability density function \( p_A(\cdot) \) in \([0, 1]\), there exists some constant \( C_0 > 0 \) such that

$$\sup_{a \in [0, 1]} p_A(a) \leq C_0 \quad \text{and} \quad \Pr(A \in \mathcal{I}) \leq C_0 |\mathcal{I}|, \quad (A.35)$$

for any interval \( \mathcal{I} \in [0, 1] \). This together with (A.34) yields that for any integer \( q \geq 1 \), \( j_1, j_2 \in \{1, \ldots, p+1\} \) and any interval \( \mathcal{I} \in [0, 1] \), we have

$$E[\mathbb{I}(A \in \mathcal{I}) |\bar{X}^{(j_1)}|^{2q} | A] = E[\mathbb{I}(A \in \mathcal{I}) E(|\bar{X}^{(j_1)}|^{2q} | A)] \leq q! \omega^{2q} E[\mathbb{I}(A \in \mathcal{I})] \leq C_0 q! \omega^{2q} |\mathcal{I}|.$$
It follows that
\[
E[\|A \in \mathcal{I}\|X_{j}^{(h)} - X_{j}^{(h)}]\|X_{i}^{(j)} - X_{i}^{(j)}\|q] - E[\|A \in \mathcal{I}\|X_{j}^{(h)} - X_{j}^{(h)}\|X_{i}^{(j)} - X_{i}^{(j)}\|q]
\leq E[\|A_{1} \in \mathcal{I}\|X_{1}^{(h)} - X_{1}^{(h)}\|X_{2}^{(j)} - X_{2}^{(j)}\|q]
\leq 2^{q-1}E[\|A_{1} \in \mathcal{I}\|X_{1}^{(h)} - X_{1}^{(h)}\|X_{2}^{(j)} - X_{2}^{(j)}\|q] + 2^{q-1}E[\|A_{2} \in \mathcal{I}\|X_{2}^{(h)} - X_{2}^{(h)}\|X_{2}^{(j)} - X_{2}^{(j)}\|q]
= 2^{q}E[\|A \in \mathcal{I}\|X_{i}^{(j)} - X_{i}^{(j)}\|\|X_{j}^{(h)} - X_{j}^{(h)}\|\leq C_{0}\|2\omega^{2}\|\mathcal{I}\|,]
\]
where the second inequality follows from Jensen's inequality and the third inequality is due to that \(|a + b|^{q} \leq 2^{q-1}|a|^{q} + 2^{q-1}|b|^{q-1}\), for any \(a, b \in \mathbb{R}\) and \(q \geq 1\).

By the Bernstein's inequality (see Lemma 2.2.11, van der Vaart and Wellner 1996), we obtain that
\[
\Pr\left(\sum_{i=1}^{n} \|A \in \mathcal{I}\|X_{i}^{(h)} - X_{i}^{(h)}\|X_{i}^{(j)} - X_{i}^{(j)}\| - nE[\|A \in \mathcal{I}\|X_{i}^{(h)} - X_{i}^{(h)}\|X_{i}^{(j)} - X_{i}^{(j)}\|] \geq t\omega^{2}\sqrt{|\mathcal{I}|n\log n}\right) \leq 2\exp\left(-\frac{t^{2}\log n}{16C_{0} + 4t(\sqrt{|\mathcal{I}|n\log n})}\right)
\]
for any \(t > 0\), any integers \(j_{1}, j_{2} \in \{1, \ldots, p + 1\}\) and any interval \(\mathcal{I} \in [0, 1]\). Set \(t = 20\sqrt{C_{0}}\), for any interval \(\mathcal{I}\) with \(|\mathcal{I}| \geq C_{0}^{-1}n^{-1}\log n\), we have
\[
t^{2}\log n \geq 4(16C_{0} + 4t(\sqrt{|\mathcal{I}|n\log n}))^{-1/2}\sqrt{\log n}.
\]
It follows from (A.36) that
\[
\Pr\left(\sum_{i=1}^{n} \|A \in \mathcal{I}\|X_{i}^{(h)} - X_{i}^{(h)}\|X_{i}^{(j)} - X_{i}^{(j)}\| - nE[\|A \in \mathcal{I}\|X_{i}^{(h)} - X_{i}^{(h)}\|X_{i}^{(j)} - X_{i}^{(j)}\|] \geq 20\omega^{2}\sqrt{C_{0}|\mathcal{I}|n\log n}\right) \leq 2n^{-4},
\]
for any integers \(j_{1}, j_{2} \in \{1, \ldots, p + 1\}\) and any interval \(\mathcal{I}\) that satisfies \(|\mathcal{I}| \geq C_{0}^{-1}n^{-1}\log n\). Notice that the number of elements in \(\mathcal{I}(m)\) is bounded by \((m + 1)^{2}\). Since \(m \asymp n\), it follows from Bonferroni's inequality that
\[
\Pr(\mathcal{A}_{1}) \geq 1 - 2(m + 1)^{2}(p + 1)^{2}n^{-4} = 1 - O(n^{-2}),
\]
where the event \( \mathcal{A}_1 \) is defined as

\[
\left\{ \left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \overline{X}_i^{(j_1)} \overline{X}_i^{(j_2)} - n \mathbb{E}(A \in \mathcal{I}) \overline{X}^{(j_1)} \overline{X}^{(j_2)} \right| \leq 20 \omega^2 \sqrt{C_0 |\mathcal{I}| n \log n} \right\}.
\]

For any symmetric matrix \( A \), we have \( \|A\|_2 \leq \|A\|_\infty \|A\|_1 = \|A\|_\infty \). Thus, under the event defined in \( \mathcal{A}_1 \), we have

\[
\left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \overline{X}_i^{(j_1)} \overline{X}_i^{(j_2)} - n \mathbb{E}(A \in \mathcal{I}) \overline{X}^{(j_1)} \overline{X}^{(j_2)} \right| \leq 20 \omega^2 (p+1) \sqrt{C_0 |\mathcal{I}| n \log n},
\]

for any \( \mathcal{I} \in \mathcal{I}(m) \) with \( |\mathcal{I}| \geq C_0^{-1} n^{-1} \log n \). Since \( \lambda_n = O(n^{-1} \log n) \), we obtain

\[
\left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \overline{X}_i^{(j_1)} \overline{X}_i^{(j_2)} - n \mathbb{E}(A \in \mathcal{I}) \overline{X}^{(j_1)} \overline{X}^{(j_2)} + n \lambda_n \mathbb{E}|\mathcal{I}| \right|_2 \leq n \lambda_n |\mathcal{I}| + \left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \overline{X}_i^{(j_1)} \overline{X}_i^{(j_2)} - n \mathbb{E}(A \in \mathcal{I}) \overline{X}^{(j_1)} \overline{X}^{(j_2)} \right|_2 \leq c \sqrt{|\mathcal{I}| n \log n},
\]

for some constant \( c > 0 \). To summarize, under the event defined in \( \mathcal{A}_1 \), we’ve shown that

\[
\left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \overline{X}_i^{(j_1)} \overline{X}_i^{(j_2)} - n \mathbb{E}(A \in \mathcal{I}) \overline{X}^{(j_1)} \overline{X}^{(j_2)} + n \lambda_n \mathbb{E}|\mathcal{I}| \right|_2 \leq c \sqrt{|\mathcal{I}| n \log n}, \quad (A.38)
\]

for any interval \( \mathcal{I} \in \mathcal{I}(m) \) with \( |\mathcal{I}| \geq C_0^{-1} n^{-1} \log n \).

Let \( \Sigma = \mathbb{E}XX^\top \). If \( \Sigma \) is singular, there exists some nonzero vector \( a \in \mathbb{R}^p \) and some \( b \in \mathbb{R} \) such that \( a^\top X = b \), almost surely. As a result, the covariance matrix of \( X \) is degenerate. Thus, we’ve reached a contraction. Therefore, \( \Sigma \) is nonsingular. There exists some constant \( \bar{c}_* > 0 \) such that

\[
\lambda_{\min}(\Sigma) \geq \bar{c}_*, \quad (A.39)
\]

By (A3), we have

\[
\Pr(A \in \mathcal{I}|X) \geq c_* |\mathcal{I}|,
\]

128
for any interval $\mathcal{I} \in [0, 1]$. This together with (A.39) implies that

$$
\lambda_{\min}\left(\mathcal{E}(A \in \mathcal{I})\mathbf{X}^T\mathbf{X}ight) = \lambda_{\min}\left(\text{EPr}(A \in \mathcal{I}|X)\mathbf{X}^T\mathbf{X}\right) \\
\geq c_{\lambda} \lambda_{\min}(\mathbf{X}\mathbf{X}^T)|\mathcal{I}| \geq c_{\lambda} |\mathcal{I}|
$$

(A.40)

For any interval $\mathcal{I}$ with $|\mathcal{I}| \geq 4c^2(c_{\lambda}\bar{c})^{-2}n^{-1}\log n$, we have

$$
c_{\lambda} \bar{c} |\mathcal{I}| - c \sqrt{|\mathcal{I}|n^{-1}\log n} \geq \frac{c_{\lambda} \bar{c} |\mathcal{I}|}{2}.
$$

In view of (A.38) and (A.40), we obtain that

$$
\lambda_{\min}\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\mathbf{X}_i^T + \lambda_n \mathcal{E}|\mathcal{I}\right) \geq \lambda_{\min}\left(\mathcal{E}(A \in \mathcal{I})\mathbf{X}^T\mathbf{X}\right) \\
\geq \frac{c_{\lambda} \bar{c} |\mathcal{I}|}{2}.
$$

(A.41)

Set $\tilde{c}_0 = \max(4c^2(c_{\lambda}\bar{c})^{-1}, C_0^{-1})$, it is immediate to see that

$$
\max_{|\mathcal{I}| \geq \tilde{c}_0 n^{-1}\log n} \eta_1(\mathcal{I}) \leq \frac{2}{c_{\lambda} \bar{c} |\mathcal{I}|^3},
$$

(A.42)

under the event defined in $\mathcal{A}_1$.

For any $\mathcal{I} \in [0, 1]$, we have

$$
\left\|\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\mathbf{X}_i^T + \lambda_n \mathcal{E}|\mathcal{I}\right)^{-1} - \left(\mathcal{E}(A \in \mathcal{I})\mathbf{X}^T\mathbf{X}\right)^{-1}\right\|_2 \\
\leq \left\|\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\mathbf{X}_i^T + \lambda_n \mathcal{E}|\mathcal{I}\right)^{-1}\right\|_2 \left\|\left(\mathcal{E}(A \in \mathcal{I})\mathbf{X}^T\mathbf{X}\right)^{-1}\right\|_2 \\
\times \left\|\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\mathbf{X}_i^T - n\mathbb{E}(A \in \mathcal{I})\mathbf{X}^T + n\lambda_n \mathcal{E}|\mathcal{I}\right\|_2
$$

This together with (A.38), (A.40) and (A.41) yields

$$
\max_{|\mathcal{I}| \geq \tilde{c}_0 n^{-1}\log n} \eta_3(\mathcal{I}) \leq \frac{2c \sqrt{n^{-1}\log n}}{c^2 \bar{c}^2 |\mathcal{I}|^{3/2}},
$$

(A.43)
under the event defined in \( \mathcal{A}_1 \).

Similar to (A.34), we can show that for any integer \( q \geq 1 \) and \( j \in \{1, \ldots, p + 1\} \),

\[
E(|X_j Y|^{q+1} | A) \leq q! \omega^2,
\]

almost surely. Specifically, set \( q = 1 \), we obtain \( E(|X_j Y| | A) \leq \omega \). By (A.35), we have that

\[
\|E \left( \sum_{j=1}^{p+1} |\mathbb{I}(A \in \mathcal{G}) X_j Y| \right)^2 \|_2 \leq \left( \sum_{j=1}^{p+1} |E \left( \mathbb{I}(A \in \mathcal{G}) X_j Y \right)|^2 \right)^{1/2} \leq \left( \sum_{j=1}^{p+1} \omega^2 |E(A \in \mathcal{G})|^2 \right)^{1/2} \leq C_0 \sqrt{p+1} \omega^2 |\mathcal{G}|.
\]

for any \( \mathcal{G} \in [0, 1] \). This implies that

\[
\max_{\mathcal{G} \in [0, 1]} \eta_4(\mathcal{G}) \leq C_0 \sqrt{p+1} \omega^2 |\mathcal{G}|.
\]

Moreover, in view of (A.36) and (A.37), we can similarly show that

\[
\Pr(\mathcal{A}_2) \geq 1 - 2(m+1)^2(p+1)n^{-4} = 1 - O(n^{-2}),
\]

where the event \( \mathcal{A}_2 \) is defined as

\[
\bigcap_{|\mathcal{G}| \geq C_0 n^{-1} \log n} \left\{ \left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{G}) X_i + n \mathbb{I}(A \in \mathcal{G}) X_j Y \right| \leq 20 \omega^2 \sqrt{C_0} |\mathcal{G}| \log n \right\}.
\]

Under the event defined in \( \mathcal{A}_2 \), we have

\[
\max_{|\mathcal{G}| \geq C_0 n^{-1} \log n} \eta_2(\mathcal{G}) \leq 20 \omega^2 \sqrt{(p+1)C_0} |\mathcal{G}| n^{-1} \log n.
\]

Combining (A.42) together with (A.43), (A.45), (A.47) yields

\[
\max_{|\mathcal{G}| \geq C_0 n^{-1} \log n} \|\hat{\theta}_\mathcal{G} - \theta_0, \mathcal{G}\|^2 = O \left( \frac{\log n}{n} \right),
\]

under the events defined in \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \). The proof is thus completed based on (A.37) and
Proofs of (A.8), (A.9) and (A.12): We first prove (A.12). By the definition of \( \theta_{0,\mathcal{S}} \), we have
\[
\| \theta_{0,\mathcal{S}} \|_2 \leq \left\| \left( \mathbb{E}XX^T \mathbb{I}(A \in \mathcal{S}) \right)^{-1} \right\|_2 \left\| \mathbb{E}X\mathbb{I}(A \in \mathcal{S}) \right\|_2.
\]
It follows from (A.35), (A.40) and (A.44) that
\[
\| \theta_{0,\mathcal{S}} \|_2 \leq \left( c_* \bar{c}_* |I| \right)^{-1} \sum_{j=1}^{n+1} \left[ \mathbb{E} \left\{ \left( \mathbb{E}X^{(j)}_I Y | A \right) \mathbb{I}(A \in \mathcal{S}) \right\} \right]^2 \leq \sqrt{p + 1} \left( c_* \bar{c}_* \right)^{-1} \bar{c}_0 \omega^2,
\]
for any \( \mathcal{S} \in [0, 1] \). Assertion (A.12) thus follows.

Consider (A.8). Since \( p \) is fixed, it suffices to show for any \( j \in \{1, \ldots, p+1\} \), the following event occurs with probability at least \( 1 - O(n^{-2}) \):
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})(Y_i - X_i^\top \theta_{0,\mathcal{S}})X_i^{(j)} \right|_2 = O \left( \frac{\sqrt{|\mathcal{S}| \log n}}{\sqrt{n}} \right). \tag{A.48}
\]
By (A.12), (A.48) can be proven in a similar manner as (A.37) and (A.46). (A.9) can be similarly proven.

Proof of (A.10): Similar to (A.8), we can show that the following event occurs with probability at least \( 1 - O(n^{-2}) \): for any \( \mathcal{S} \in \mathcal{I}(m) \) such that \( |\mathcal{S}| \geq \bar{c}_0 n^{-1} \log n \),
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[X_i^\top (\theta_0(A_i) - \theta_{0,\mathcal{S}})]^2 - \mathbb{E}(A \in \mathcal{S})[X^\top (\theta_0(A) - \theta_{0,\mathcal{S}})]^2 \right|_2 = O \left( \frac{\sqrt{|\mathcal{S}| \log n}}{\sqrt{n}} \right).
\]
Notice that
\[
\mathbb{E}(A \in \mathcal{S})[X^\top (\theta_0(A) - \theta_{0,\mathcal{S}})]^2 = \int_{\mathcal{S}} [X^\top (\theta_0(a) - \theta_{0,\mathcal{S}})]^2 \pi(a | X) da \\
\geq c_* \int_{\mathcal{S}} [X^\top (\theta_0(a) - \theta_{0,\mathcal{S}})]^2 da = c_* \int_{\mathcal{S}} \{\theta_0(a) - \theta_{0,\mathcal{S}}\}^\top \Sigma \{\theta_0(a) - \theta_{0,\mathcal{S}}\} da \\
\geq c_* \lambda_{\min}(\Sigma) \int_{\mathcal{S}} \|\theta_0(a) - \theta_{0,\mathcal{S}}\|_2^2 da + c_* \mathcal{E}_* \int_{\mathcal{S}} \|\theta_0(a) - \theta_{0,\mathcal{S}}\|_2^2 da, \tag{A.49}
\]
where the first inequality is due to Condition (A3) and the last inequality is due to (A.39).
It follows that
\[ \frac{1}{n} \sum_{i=1}^{n} I(A_i \in \mathcal{I}) \{ \theta_0(A_i) - \theta_{0,\mathcal{I}} \}^2 \geq c_\ast \bar{c} \int_{\mathcal{I}} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2^2 \, da - O \left( \frac{\sqrt{\mathcal{I} \log n}}{\sqrt{n}} \right), \]
for any $\mathcal{I} \in \mathcal{I}(m)$ such that $|\mathcal{I}| \geq \bar{c}_0 n^{-1} \log n$, with probability at least $1 - O(n^{-2})$. This completes the proof.

**Proof of (A.11):** Similar to (A.8), we can show that the following event occurs with probability at least $1 - O(n^{-2})$: for any $\mathcal{I} \in \mathcal{I}(m)$ such that $|\mathcal{I}| \geq \bar{c}_0 n^{-1} \log n$,
\[
\left| \frac{1}{n} \sum_{i=1}^{n} I(A_i \in \mathcal{I}) \{ |Y_i|^2 + \|X_i\|^2 \} - \mathbb{E}(A \in \mathcal{I}) (Y^2 + \|X\|^2) \right| = O \left( \frac{\sqrt{|\mathcal{I}| \log n}}{\sqrt{n}} \right). \tag{A.50}
\]

By (A.35) and (A.33), we have
\[
\mathbb{E}(A \in \mathcal{I}) \|X\|^2 \leq \sum_{j=1}^{p+1} \mathbb{E}(A \in \mathcal{I}) |X^{(j)}|^2 \leq (p + 1) C_0 \omega^2 |\mathcal{I}|.
\]

Similarly, we can show
\[
\mathbb{E}(A \in \mathcal{I}) Y^2 \leq C_0 \omega^2 |\mathcal{I}|,
\]
and thus
\[
\mathbb{E}(A \in \mathcal{I}) (Y^2 + \|X\|^2) \leq (p + 2) C_0 \omega^2 |\mathcal{I}|.
\]

This together with (A.50) yields (A.11).

**A.2.3 Proof of Lemma 2**

We first prove (A.13). By (A.12), we have $\sup_{a \in [0,1]} \| \theta_0(a) \|_2 \leq c_0$ and hence
\[
\sup_{a \in [0,1]} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2 \leq 2c_0. \tag{A.51}
\]

Similarly, we can show that for any integer $q \geq 1$,
\[
\mathbb{E}(|Y|^{2q} | A) \leq q! \omega^{2q}. \tag{A.52}
\]
For any \( \mathcal{S} \subseteq \mathcal{J}(m) \) and integer \( q \geq 2 \), it follows from (A.44), (A.51) and (A.52) that

\[
E\left( [Y X^T \{ \theta_0(A) - \theta_{0,\mathcal{S}} \}]^q | A \right) \leq \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^q \| Y \| \| X \|^q | A \] 
\[
\leq \frac{1}{2} \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^2 E\left( |Y|^{2q} + \sum_{j=1}^{p+1} (X_j)^2 \right) \leq \frac{1}{2} \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^q q! \omega^{2q}
\]
\[
+ \frac{1}{2} \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^2 (p + 1)^q \sum_{j=1}^{p+1} E(|X_j|^2 | A) \leq \frac{q! \omega^{2q}}{2} (1 + (p + 1)^q) \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^q
\]
\[
\leq q! \omega^{2q} (p + 1)^q \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^q \leq q! \omega^{2q} (p + 1)^q (2c_0)^{q-2} \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^2.
\]

Similarly, we can show

\[
E\left( [\bar{X}^T \theta_0(A)] X^T \{ \theta_0(A) - \theta_{0,\mathcal{S}} \}]^q | A \right) \leq q! \omega^{2q} (p + 1)^q 2^{q-2} c_0^{q-2} \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^2.
\]

This together with (A.53) yields that for any integer \( q \geq 2 \), \( \mathcal{S} \subseteq [0, 1] \), we have

\[
E\left( [Y - \bar{X}^T \theta_0(A)] X^T \{ \theta_0(A) - \theta_{0,\mathcal{S}} \}]^q | A \right) \leq q! c^q \| \theta_0(A) - \theta_{0,\mathcal{S}} \|^2,
\]

for some constant \( c > 0 \). Combining (A.35) together with (A.54), we obtain that for any integer \( q \geq 2 \), \( \mathcal{S} \subseteq [0, 1] \),

\[
E[\mathbb{I}(A \in \mathcal{S}) \{Y - \bar{X}^T \theta_0(A)] X^T \{ \theta_0(A) - \theta_{0,\mathcal{S}} \}] \leq C_0 q! c^q \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0,\mathcal{S}} \|^2 p_\lambda(a) d a
\]
\[
\leq C_0 q! c^q \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0,\mathcal{S}} \|^2 d a.
\]

Applying the Bernstein’s inequality (using similar arguments in (A.36) and (A.37)), we can show that with probability at least \( 1 - O(n^{-2}) \), we have for any interval \( \mathcal{S} \) that satisfies \( \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0,\mathcal{S}} \|^2 d a \geq (C_0)^{-1} n^{-1} \log n \) and \( \mathcal{S} \in \mathcal{J}(m) \),

\[
\sum_{i=1}^n \mathbb{I}(A_i \in \mathcal{S}) \{Y_i - \bar{X}_i^T \theta_0(A_i)] X_i^T \{ \theta_0(A_i) - \theta_{0,\mathcal{S}} \}] \leq O(1) \sqrt{n \log n} \left( \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0,\mathcal{S}} \|^2 d a \right)^{1/2},
\]

where \( O(1) \) denotes some positive constant. This proves (A.13).

Similarly, we can show that with probability at least \( 1 - O(n^{-2}) \), there exists some constant \( C > 0 \) such that for any interval \( \mathcal{S} \) that satisfies \( \int_{\mathcal{S}} \| \theta_0(a) - \theta_{0,\mathcal{S}} \|^2 d a \geq (C_0)^{-1} n^{-1} \log n \) and
\( I \in \mathcal{I}(m) \), we have
\[
\frac{\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \{\theta_0(A_i) - \theta_{0,r}\}^2}{n} - n \mathbb{E} [A \in \mathcal{I}] \{\theta_0(A) - \theta_{0,r}\}^2 \leq O(1) \sqrt{n \log n} \left( \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,r}\|_2^2 da \right)^{1/2},
\]
for some positive constant \( O(1) \). This together with (A.49) yields (A.14).

### A.2.4 Proof of Lemma 3

Consider the following three categories of intervals.

**Category 1:** Suppose \( i_1 \) and \( i_2 \) satisfy \( \tau_{0,k-1} \leq i_1 \leq i_2 \leq \tau_{0,k} \) for some integer \( k \) such that \( 1 \leq k \leq K \). Then apparently, we have \( \theta_{0,r} = \theta_0(a), \forall a \in \mathcal{I} \), and hence \( \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,r}\|_2^2 da = 0 \).

The assertion \( \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,r}\|_2^2 da \leq c_n \) is thus automatically satisfied.

**Category 2:** Suppose there exists some integer \( k \) such that \( 2 \leq k \leq K \) and \( i_1, i_2 \) satisfy \( \tau_{0,k-2} \leq i_1 < \tau_{0,k-1} < i_2 \leq \tau_{0,k} \). Assume we have
\[
\min_{j \in \{1, 2\}} |i_j - \tau_{0,k-1}| \geq \frac{3}{\kappa_0} c_n.
\]
where
\[
\kappa_0 \equiv \min_{\mathcal{I}_1, \mathcal{I}_2 \in \mathcal{P}_0} \|\theta_{0,\mathcal{I}_1} - \theta_{0,\mathcal{I}_2}\|_2 > 0.
\]

Since \( c_n \to 0 \), for sufficiently large \( n \), we have \( \tau_{0,k} > \tau_{0,k-1} + 3\kappa_0^{-2} c_n \) and \( \tau_{0,k-2} + 3\kappa_0^{-2} c_n < \tau_{0,k-1} \). Then, we have
\[
\int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,r}\|_2^2 da \geq \min_{\theta \in \mathbb{R}^{p+1}} \int_{\tau_{0,k-1} - 3\kappa_0^{-2} c_n}^{\tau_{0,k-1} + 3\kappa_0^{-2} c_n} \|\theta - \theta_0(a)\|_2^2 da \geq \frac{6}{\kappa_0^2} c_n \min_{\theta \in \mathbb{R}^{p+1}} (\|\theta - \theta_0|_{\tau_{0,k-2}, \tau_{0,k-1}}\|_2^2, \|\theta - \theta_0|_{\tau_{0,k-1}, \tau_{0,k}}\|_2^2) \geq \frac{6}{\kappa_0^2} c_n \kappa_0^{-2} > c_n.
\]
This violates the assertion that \( \int_{\mathcal{I}} \|\theta_0(a) - \theta_{0,r}\|_2^2 da \leq c_n \). We've thus reached a contradiction.
As a result, we have

\[
\min_{j \in \{1, 2\}} |i_j - \tau_{0,k-1}| \leq \frac{3}{\kappa_0^2} c_n.
\]

**Category 3:** Suppose there exists some integer \(k\) such that \(3 \leq k \leq K\) and \(i_1, i_2\) satisfy \(\tau_{0,k-3} \leq i_1 < \tau_{0,k-2} < \tau_{0,k-1} < i_2 \leq \tau_{0,k}\). Assume we have

\[
|i_1 - \tau_{0,k-2}| \geq \frac{3}{\kappa_0^2} c_n.
\]

Then for sufficiently large \(n\), we have

\[
\int_{\mathcal{J}} \|\theta_0(a) - \theta_0,0\|_2^2 da \geq \min_{\theta \in \mathbb{R}^{p+1}} \int_{\tau_{0,k-2}^2}^{\tau_{0,k-2}+3\kappa_0^{-2} c_n} \|\theta - \theta_0(a)\|_2^2 da \\
\geq \frac{6}{\kappa_0^{-2}} c_n \min_{\theta \in \mathbb{R}^{p+1}} \left( \|\theta - \theta_0,0,\tau_{0,k-2},\tau_{0,k-1}\|_2^2, \|\theta - \theta_0,0,\tau_{0,k-1},\tau_{0,k}\|_2^2 \right) \geq \frac{6}{\kappa_0^{-2}} c_n \frac{\kappa_0^{-2}}{4} > c_n.
\]

This violates the assertion that \(\int_{\mathcal{J}} \|\theta_0(a) - \theta_0,0\|_2^2 da \leq c_n\). We’ve thus reached a contradiction. As a result, we have \(|i_1 - \tau_{0,k-2}| \leq 3\kappa_0^{-2} c_n\). Similarly, we can show \(|i_2 - \tau_{0,k-1}| \leq 3\kappa_0^{-2} c_n\).

Therefore, we obtain

\[
\max_{j \in \{1, 2\}} |i_j - \tau_{0,k-3+j}| \leq \frac{3}{\kappa_0^2} c_n.
\]

If \(\mathcal{J}\) belongs to none of these categories, then there exists some integer \(k\) such that \(2 \leq k \leq K\) and \(i_1, i_2\) satisfy \(i_1 \leq \tau_{0,k-2}\) and \(i_2 \geq \tau_{0,k}\). Using similar arguments, we can show that

\[
\int_{\mathcal{J}} \|\theta_0(a) - \theta_0,0\|_2^2 da \geq \int_{\tau_{0,k}}^{\tau_{0,k-2}} \|\theta_0(a) - \theta_0,0\|_2^2 da \geq \frac{\kappa_0^2}{4} \min_{\mathcal{J} \in \mathcal{J}} |\mathcal{J}|.
\]

For sufficiently large \(n\), this violates the assertion that \(\int_{\mathcal{J}} \|\theta_0(a) - \theta_0,0\|_2^2 da \leq c_n\). We’ve thus reached a contradiction. Therefore, we shall have \(\tau_{0,k-2} \leq i_1 < i_2 \leq \tau_{0,k}\). This completes the first part of the proof.

We now show (A.15). Take \(c_n = \bar{c}_1 n^{-1} \log n\) and consider any interval \(\mathcal{J} \in \mathcal{J}(m)\) that satisfies \(\int_{\mathcal{J}} \|\theta_0(a) - \theta_0,0\|_2^2 da \leq \bar{c}_1 n^{-1} \log n\).
If $\mathcal{J}$ belongs to Category 1, then $\theta_0(a) = \theta_{0, \mathcal{J}}$ for any $a \in \mathcal{J}$. As a result, we have

$$\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \{ \theta_0(A_i) - \theta_{0, \mathcal{J}} \} = 0. $$

If $\mathcal{J}$ belongs to Category 2, then there exists some integer $k$ such that $2 \leq k \leq K$ and $i_1, i_2$ satisfy $\tau_{0,k-2} \leq i_1 < \tau_{0,k-1} < i_2 \leq \tau_{0,k}$. Thus, we have

$$\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \{ \theta_0(A_i) - \theta_{0, \mathcal{J}} \} = \sum_{i=1}^{n} \mathbb{I}(A_i \in [i_1, \tau_{0,k-1}]) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \{ \theta_0(A_i) - \theta_{0, \mathcal{J}} \} \zeta_1 + \sum_{i=1}^{n} \mathbb{I}(A_i \in [\tau_{0,k-1}, i_2]) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \{ \theta_0(A_i) - \theta_{0, \mathcal{J}} \} \zeta_2.$$

Notice that we’ve shown

$$\min_{j \in \{1, 2\}} |i_j - \tau_{0,k-1}| \leq \frac{3\bar{c}_1}{k_0^2} n^{-1} \log n. $$

Without loss of generality, suppose $|i_1 - \tau_{0,k-1}| \leq 3\bar{c}_1 k_0^{-2} n^{-1} \log n$. Using similar arguments in (A.22) and (A.23), we can show that $\zeta_1 = O(\log n)$, with probability at least $1 - O(n^{-2})$.

As for $\zeta_2$, consider intervals of the form $[\tau_{0,j}, (m+1)i]^{-1}i$ for $j = 0, 1, \ldots, K-1, i = 1, \ldots, m+1$. Denoted by $\mathcal{J}(m)$ the set consisting of all such intervals. Similar to Lemma 1, we can show that the following event occurs with probability at least $1 - O(n^{-2})$:

$$\left\| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{J}) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \right\|_2 = O(\sqrt{n|\mathcal{J}||\log n}), \quad \text{(A.55)}$$

for any $\mathcal{J} \in \mathcal{J}(m)$ with $|\mathcal{J}| \geq cn^{-1} \log n$ for some constant $c > 0$. Suppose $i_2 - \tau_{0,k-1} \geq cn^{-1} \log n$. Under the event defined in (A.55), it follows that

$$\left\| \sum_{i=1}^{n} \mathbb{I}(A_i \in [\tau_{0,k-1}, i_2]) \{ Y_i - \overline{X}_i^\top \theta_0(A_i) \} \overline{X}_i \right\|_2 = O(\sqrt{n|\mathcal{J}||\log n}), \quad \text{(A.56)}$$

Since $\int_{\mathcal{J}} \| \theta_0(a) - \theta_{0, \mathcal{J}} \|^2 d a \leq \bar{c}_1 n^{-1} \log n$, we have $\int_{\tau_{0,k-1}}^{i_2} \| \theta_0(a) - \theta_{0, \mathcal{J}} \|^2 d a \leq \bar{c}_1 n^{-1} \log n$, and
hence \((i_2 - \tau_{0,k-1})||\theta_0(a) - \theta_{0,\mathcal{F}}||_2^2 \leq \bar{c}_1 n^{-1} \log n\), for any \(a \in [\tau_{0,k-1}, i_2]\). This together with (A.56) yields that

\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in [\tau_{0,k-1}, i_2])\{Y_i - X_i^{\top} \theta_0(A_i)\} \bar{X}_i \{\theta_0(A_i) - \theta_{0,\mathcal{F}}\} \\
\leq \sum_{i=1}^{n} \mathbb{I}(A_i \in [\tau_{0,k-1}, i_2])\{Y_i - X_i^{\top} \theta_0(A_i)\} \bar{X}_i \ ||\theta_0(\tau_{0,k-1}) - \theta_{0,\mathcal{F}}||_2 = O(\log n),
\]

and hence \(\zeta_2 = O(\log n)\). When \(i_2 - \tau_{0,k-1} \leq cn^{-1} \log n\), using similar arguments in (A.22) and (A.23), we can show that \(\zeta_2 = O(\log n)\), with probability at least \(1 - O(n^{-2})\). Thus, we've shown that with probability at least \(1 - O(n^{-2})\), for any interval \(\mathcal{I}\) that belongs to the Category 2 with \(\int_{\mathcal{I}} ||\theta_0(a) - \theta_{0,\mathcal{F}}||_2^2 \leq \bar{c}_1 n^{-1} \log n\), we have

\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\{Y_i - X_i^{\top} \theta_0(A_i)\} \bar{X}_i \{\theta_0(A_i) - \theta_{0,\mathcal{F}}\} = O(\log n).
\]

Similarly, one can show that with probability at least \(1 - O(n^{-2})\), for any interval \(\mathcal{I}\) that belongs to the Category 3 with \(\int_{\mathcal{I}} ||\theta_0(a) - \theta_{0,\mathcal{F}}||_2^2 \leq \bar{c}_1 n^{-1} \log n\), we have

\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\{Y_i - X_i^{\top} \theta_0(A_i)\} \bar{X}_i \{\theta_0(A_i) - \theta_{0,\mathcal{F}}\} = O(\log n).
\]

The proof is thus completed.

### A.2.5 Proof of Lemma 4

Consider a given interval \(\mathcal{I} \in \mathcal{\bar{F}}\). Suppose \(|\mathcal{I}| < \bar{c}_3 \gamma_n\). The value of the constant \(\bar{c}_3\) will be determined later. Then, for sufficiently large \(n\), we can find some interval \(\mathcal{I}' \in \mathcal{I}(m) \cap \mathcal{\bar{F}}\) that is adjacent to \(\mathcal{I}\). Thus, we have \(\mathcal{I} \cup \mathcal{I}' \in \mathcal{I}(m)\), and hence

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})\{Y_i - X_i^{\top} \hat{\theta}_{\mathcal{I}}\}^2 + \lambda_n |\mathcal{I}| ||\hat{\theta}_{\mathcal{I}}||_2^2 + \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}')\{Y_i - X_i^{\top} \hat{\theta}_{\mathcal{I}'\mathcal{I}}\}^2 + \lambda_n |\mathcal{I} \cup \mathcal{I}'| ||\hat{\theta}_{\mathcal{I} \cup \mathcal{I}'}||_2^2 - \gamma_n \quad (A.57)
\]

\[
+ \lambda_n |\mathcal{I}'| ||\hat{\theta}_{\mathcal{I}'}||_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}')\{Y_i - X_i^{\top} \hat{\theta}_{\mathcal{I} \cup \mathcal{I}'}\}^2 + \lambda_n |\mathcal{I} \cup \mathcal{I}'| ||\hat{\theta}_{\mathcal{I} \cup \mathcal{I}'}||_2^2 - \gamma_n.
\]
Notice that the left-hand-side of the above expression is nonnegative. It follows that

\[ \gamma_n \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') (Y_i - \overline{X}_i \theta_{\mathcal{I} \cup \mathcal{I}'})^2 + \lambda_n |\mathcal{I} \cup \mathcal{I}'| \|\bar{\theta}_{\mathcal{I} \cup \mathcal{I}'}\|^2. \]

By definition, we have

\[ \bar{\theta}_{\mathcal{I} \cup \mathcal{I}'} = \arg \min_{\theta \in \mathbb{R}^{p+1}} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') (Y_i - \overline{X}_i \theta)^2 + \lambda_n |\mathcal{I} \cup \mathcal{I}'| \|\theta\|^2 \right). \quad (A.58) \]

Therefore, we obtain that

\[
\gamma_n \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') (Y_i - \overline{X}_i 0_{p+1})^2 + \lambda_n |\mathcal{I} \cup \mathcal{I}'| \|0_{p+1}\|^2
= \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') Y_i^2.
\quad (A.59)
\]

Suppose

\[ |\mathcal{I} \cup \mathcal{I}'| \leq \frac{\gamma_n}{8c_0}, \quad (A.60) \]

where the constant \( c_0 \) is defined in Lemma 1.

Since \( \gamma_n \gg n^{-1} \) and \( m \asymp n \), we can find some interval \( \mathcal{I}^* \in \mathcal{I}(m) \) that covers \( \mathcal{I} \cup \mathcal{I}' \) and satisfies \((8c_0)^{-1} \gamma_n \leq |\mathcal{I}^*| \leq (4c_0)^{-1} \gamma_n \). Under the event defined in (A.11), it follows from the condition \( \gamma_n \gg n^{-1} \log n \) that

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') Y_i^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}^*) Y_i^2 \leq c_0 \left( \frac{\sqrt{(4c_0)^{-1} \gamma_n \log n}}{\sqrt{n}} + (4c_0)^{-1} \gamma_n \right)
\leq 2c_0(4c_0)^{-1} \gamma_n = \frac{\gamma_n}{2},
\]

for sufficiently large \( n \). This apparently violates the results in (A.59). Thus, Assertion (A.60) doesn't hold. Therefore, we obtain that

\[ |\mathcal{I} \cup \mathcal{I}'| \geq \frac{\gamma_n}{8c_0}, \quad (A.61) \]

with probability at least \( 1 - O(n^{-2}) \).

Suppose the constant \( \tilde{c}_3 \) satisfies \( \tilde{c}_3 \leq (16c_0)^{-1} \). Under the event defined in (A.61), we have
\[ |\mathcal{S}'| \geq \gamma_n (16c_0)^{-1} \]. By (A.7), we have with probability at least \( 1 - O(n^{-2}) \) that \( \| \hat{\theta}_{\mathcal{S}'} - \theta_{0,\mathcal{S}'} \|_2 \leq c_0 \sqrt{n^{-1} \log n} |\mathcal{S}'|^{-1/2} \leq 4c_0^{3/2} n^{-1} \log n \gamma_n^{-1} \ll 1 \). By (A.12), we have with probability at least \( 1 - O(n^{-2}) \) that \( \| \hat{\theta}_{\mathcal{S}'} \|_2 \leq 2c_0 \), (A.62) for sufficiently large \( n \).

In addition, it follows from (A.58) that
\[
\frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S} \cup \mathcal{S}') \left( Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S} \cup \mathcal{S}'} \right)^2 + \lambda_n |\mathcal{S} \cup \mathcal{S}'| \| \hat{\theta}_{\mathcal{S} \cup \mathcal{S}'} \|_2^2 \\
\leq \frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S} \cup \mathcal{S}') \left( Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S}'}, \right)^2 + \lambda_n |\mathcal{S} \cup \mathcal{S}'| \| \hat{\theta}_{\mathcal{S}'} \|_2^2.
\]

By (A.57), this further implies that
\[
\frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) \left( Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S}} \right)^2 + \lambda_n |\mathcal{S}| \| \hat{\theta}_{\mathcal{S}} \|_2^2 \\
\leq \frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) \left( Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S}}, \right)^2 + \lambda_n |\mathcal{S}| \| \hat{\theta}_{\mathcal{S}} \|_2^2 - \gamma_n,
\]
and hence
\[
\gamma_n \leq \frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) \left( Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S}} \right)^2 + \lambda_n |\mathcal{S}| \| \hat{\theta}_{\mathcal{S}} \|_2^2.
\]

By (A.62) and the conditions that \( \lambda_n = O(n^{-1} \log n) \), \( \gamma_n \gg n^{-1} \log n \), we have for sufficiently large \( n \),
\[
\frac{\gamma_n}{2} \leq \frac{1}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) (Y_i - \overline{X}_i \top \hat{\theta}_{\mathcal{S}})^2.
\]

It thus follows from Cauchy-Schwarz inequality and (A.62) that
\[
\frac{\gamma_n}{2} \leq \frac{2}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) (Y_i^2 + \| \overline{X}_i \|_2^2 \| \hat{\theta}_{\mathcal{S}} \|_2^2) \leq \frac{2(1 + 4c_0^2)}{n} \sum_{i=1}^{n} 1(A_i \in \mathcal{S}) (Y_i^2 + \| \overline{X}_i \|_2^2).
\]

Using similar arguments in showing (A.61), we obtain that
\[
|\mathcal{S}| \geq \frac{\gamma_n}{32(1 + 4c_0^2)c_0}.
\]
with probability at least $1 - O(n^{-2})$. Set $\bar{c}_3 = 32^{-1}(1 + 4c_0^2)^{-1}c_0^{-1}$, this violates the assumption that $|\mathcal{I}| < \bar{c}_3\gamma_n$. Thus, with probability at least $1 - O(n^{-2})$, we obtain that $|\mathcal{I}| \geq \bar{c}_3\gamma_n$, for any $\mathcal{I} \in \mathcal{P}$. The proof is hence completed.

### A.2.6 Proof of Theorem 2

Let $\{\hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_{K-1}\}$ be the set of change points in $J(\mathcal{P})$. Under the events defined in Theorem 1, we have $\hat{K} = K$, and

$$\max_{k \in \{1, \ldots, K-1\}} |\hat{\tau}_k - \tau_{0,k}| \leq c n^{-1} \log n, \quad (A.63)$$

for some constant $c > 0$. Set $\hat{\tau}_0 = 0$ and $\hat{\tau}_K = 1$.

Under the event defined in (A.63), we have for sufficiently large $n$ that

$$\hat{\tau}_k - \hat{\tau}_{k-1} \geq \delta_{\min}, \quad \forall k \in \{1, \ldots, K\}. \quad (A.64)$$

Since $\pi^*$ satisfies $\sup_{\mathcal{I} \subseteq [0,1]} \sup_{a \in \mathcal{I}, x \in \mathcal{X}} |\mathcal{I}| \pi^*(a; x, \mathcal{I}) > 1$, there exists some constant $\bar{c}_4 > 0$ such that $\pi^*(a; x, \bar{d}(x)) \leq \bar{c}_4|\bar{d}(x)|^{-1}$ for all $a$ and $x$. This together with (A.64) yields that

$$\pi^*(a; x, \bar{d}(x)) \leq \bar{c}_4 \delta_{\min}^{-1}, \quad \forall a \in [0,1], x \in \mathcal{X}. \quad (A.65)$$

The rest of our proof is divided into three parts. In the first part, we show that there exists some constant $C > 0$ such that

$$\|\hat{\theta}_{[\hat{\tau}_{k-1}, \hat{\tau}_k]} - \theta_{[\tau_{0,k-1}, \tau_{0,k}]}\|_2 \leq \frac{C \log n}{n}, \quad \forall k \in \{1, \ldots, K\}, \quad (A.66)$$

with probability at least $1 - O(n^{-2})$. Using similar arguments in Lemma 1, we can show that there exists some constant $c_3 > 0$ such that the following events occur with probability at least $1 - O(n^{-2})$:

$$\|\hat{\theta}_{[\tau_{0,k-1}, \tau_{0,k}]} - \theta_{[\tau_{0,k-1}, \tau_{0,k}]}\|_2 \leq \frac{c_3 \sqrt{\log n}}{\sqrt{n\delta_{\min}}}, \quad \forall k \in \{1, \ldots, K\}. \quad (A.67)$$

This together with (A.66) implies that

$$\|\hat{\theta}_{[\hat{\tau}_{k-1}, \hat{\tau}_k]} - \theta_{[\tau_{0,k-1}, \tau_{0,k}]}\|_2 \leq \frac{2c_3 \sqrt{\log n}}{\sqrt{n\delta_{\min}}}, \quad \forall k \in \{1, \ldots, K\}, \quad (A.67)$$
for sufficiently large $n$, with probability at least $1 - O(n^{-2})$.

In the second part, we define an integer-valued function $\hat{K}(x)$ as follows. We set $\hat{K}(x) = k$ if $\hat{d}(x) = [\hat{\tau}_{k-1}, \hat{\tau}_k)$ for some integer $k$ such that $1 \leq k \leq K - 1$, and set $\hat{K}(x) = K$ if $\hat{d}(x) = [\hat{\tau}_{K-1}, 1]$. By the definition of $\hat{\theta}_\sigma$ and $\theta_{0, \sigma}$, we have almost surely $\hat{\theta}_{[\hat{\tau}_{K-1}, 1]} = \hat{\theta}_{[\hat{\tau}_{K-1}, 1]}$ and $\theta_{0, [\tau_{0, K-1}, 1]} = \theta_{0, [\tau_{0, K-1}, 1]}$. It is immediate to see that

$$
\hat{K}(x) = \text{sarg max}_{k \in [1, \ldots, K]} x^\top \hat{\theta}_{[\hat{\tau}_{k-1}, \hat{\tau}_k]},
$$

(A.68)

where sargmax denotes the smallest maximizer when the argmax is not unique. In Part 2, we focus on proving

$$
V^{\pi^*}(\hat{d}) \geq E\left(\sum_{i=1}^{n} I(\tau_{0, k-1} \leq A_i < \tau_{0, k}) X_i X_i^\top + \lambda_n (\tau_{0, k} - \tau_{0, k-1}) \mathbb{E}_{p+1}\right) - O(1)n^{-1} \log n,
$$

(A.69)

with probability at least $1 - O(n^{-2})$, where $O(1)$ denotes some positive constant.

In the last part, we provide an upper bound for

$$
V^{opr} - E\left(\sum_{i=1}^{n} I(\tau_{0, k-1} \leq A_i < \tau_{0, k}) X_i X_i^\top\right).
$$

This together with (A.69) yields the desired results.

**Proof of Part 1:** Let $\Delta_k = [\tau_{k-1}, \tau_k) \cup [\tau_{0, k-1}, \tau_{0, k}) \cup [\tau_{k-1}^c, \tau_k^c] \cup [\tau_{0, k-1}^c, \tau_{0, k}^c]$. With some calculations, we can show that

$$
\|\hat{\theta}_{[\tau_{k-1}, \tau_k]} - \theta_{0, [\tau_{k-1}, \tau_k]}\|_2 \leq \zeta_1(k)\zeta_2(k) + \zeta_3(k)\zeta_4(k),
$$

where

$$
\zeta_1(k) = \left\| \left( \frac{1}{n} \sum_{i=1}^{n} I(\tau_{0, k-1} \leq A_i < \tau_{0, k}) X_i X_i^\top + \lambda_n (\tau_{0, k} - \tau_{0, k-1}) \mathbb{E}_{p+1} \right)^{-1} \right\|_2,
$$

$$
\zeta_2(k) = \left\| \frac{1}{n} \sum_{i=1}^{n} I(A_i \in \Delta_k) X_i Y_i \right\|_2,
$$

$$
\zeta_3(k) = \left\| \frac{1}{n} \sum_{i=1}^{n} I(\tau_{0, k-1} \leq A_i < \tau_{0, k}) X_i Y_i \right\|_2,
$$

$$
\zeta_4(k) = \left\| \left( \frac{1}{n} \sum_{i=1}^{n} I(\tau_{0, k-1} \leq A_i < \tau_{0, k}) X_i X_i^\top + \lambda_n (\tau_{0, k} - \tau_{0, k-1}) \mathbb{E}_{p+1} \right)^{-1} \right\|_2.
$$
Similar to (A.42), we can show with probability at least $1 - O(n^{-2})$ that

$$\max_{k \in \{1, \ldots, K\}} \zeta_1(k) = O(1) \quad \text{and} \quad \max_{k \in \{1, \ldots, K\}} \zeta_3(k) = O(1), \quad (A.70)$$

where

$$\zeta_5(k) = \left\| \left( \frac{1}{n} \sum_{i=1}^{n} I(\tau_{k-1} \leq A_i < \hat{\tau}_k) \bar{X}_i \bar{X}_i^\top + \lambda_n (\tau_{k-1} - \hat{\tau}_k) E_{p+1} \right)^{-1} \right\|_2.$$

Under the event defined in (A.63), the Lebesgue measure of $\Delta_k$ is uniformly bounded by $2c n^{-1} \log n$, for any $k \in \{1, \ldots, K\}$. Using similar arguments in (A.22) and (A.23), we can show with probability at least $1 - O(n^{-2})$ that

$$\max_{k \in \{1, \ldots, K\}} \zeta_2(k) = O(n^{-1} \log n). \quad (A.71)$$

Similar to (A.45), we can show with probability at least $1 - O(n^{-2})$ that

$$\max_{k \in \{1, \ldots, K\}} \zeta_3(k) = O(1). \quad (A.72)$$

Notice that $\zeta_4(k)$ can be upper bounded by

$$\zeta_4(k) \leq \zeta_1(k) \zeta_5(k) \left\| \frac{1}{n} \sum_{i=1}^{n} I(\tau_{0,k-1} \leq A_i < \tau_{0,k}) \bar{X}_i \bar{X}_i^\top + \lambda_n (\tau_{0,k} - \tau_{0,k-1}) E_{p+1} \right\|_2
- \frac{1}{n} \sum_{i=1}^{n} I(\tau_{k-1} \leq A_i < \hat{\tau}_k) \bar{X}_i \bar{X}_i^\top - \lambda_n (\hat{\tau}_k - \tau_{k-1}) E_{p+1} \right\|_2
\leq \zeta_1(k) \zeta_5(k) \left\| \frac{1}{n} \sum_{i=1}^{n} I(A_i \in \Delta_k) \bar{X}_i \bar{X}_i^\top + \lambda_n (\tau_{0,k} - \tau_{0,k-1} - \hat{\tau}_k + \hat{\tau}_k - \tau_{k-1}) E_{p+1} \right\|_2.$$

Under the condition $\lambda_n = O(n^{-1} \log n)$, using similar arguments in (A.22) and (A.23), we can show that with probability at least $1 - O(n^{-2})$, the absolute value of each element in the matrix

$$\frac{1}{n} \sum_{i=1}^{n} I(A_i \in \Delta_k) \bar{X}_i \bar{X}_i^\top + \lambda_n (\tau_{0,k} - \tau_{0,k-1} - \hat{\tau}_k + \hat{\tau}_k - \tau_{k-1}) E_{p+1}$$
is upper bounded by $O(n^{-1} \log n)$, uniformly for any $k \in \{1, \ldots, K\}$. It follows that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \Delta_k) \bar{X}_i \bar{X}_i^\top + \lambda_n (\tau_{0,k} - \tau_{0,k-1} - \bar{\tau}_k + \bar{\tau}_{k-1}) \mathbb{E}_{p+1} \right\|_2 = O(n^{-1} \log n).$$

In view of (A.70), we obtain that

$$\max_{k \in \{1, \ldots, K\}} \zeta_4(k) = O(n^{-1} \log n), \quad (A.73)$$

with probability at least $1 - O(n^{-2})$. Combining (A.70)-(A.73) yields (A.66).

**Proof of Part 2:** It follows from Condition (A4) and the definition of the conditional Orlicz norm that

$$E \left\{ \exp \left( \frac{|X(j)|^2}{\omega^2} \right) \right\} = E \left[ E \left\{ \exp \left( \frac{|X(j)|^2}{\omega^2} \right) \right| A \right] \leq 2,$$

for any $j \in \{1, \ldots, p\}$. Without loss of generality, suppose $\omega \geq \log^{-1/2} 2$. Then, we have

$$E \left\{ \exp \left( \frac{\bar{X}(j)^2}{\omega^2} \right) \right\} = E \left[ E \left\{ \exp \left( \frac{\bar{X}(j)^2}{\omega^2} \right) \right| A \right] \leq 2,$$

for any $j \in \{1, \ldots, p+1\}$. As a result, it follows from Bonferroni’s inequality and Markov’s inequality that

$$\Pr \left( \|\bar{X}\|_2 > \omega \sqrt{2(p+1) \log n} \right) \leq \sum_{j=1}^{p+1} \Pr(|\bar{X}(j)| > \omega \sqrt{2 \log n})$$

$$\leq \sum_{j=1}^{p+1} E \left\{ \exp \left( \frac{|\bar{X}(j)|^2}{\omega^2} \right) \right\} / \exp \left( \frac{2 \omega^2 \log n}{\omega^2} \right) \leq \frac{2(p+1)}{n^2}.$$

Thus, we obtain

$$\Pr(\mathcal{A}^*) \geq 1 - \frac{2(p+1)}{n^2}, \quad (A.74)$$

where

$$\mathcal{A}^* = \{\|\bar{X}\|_2 \leq \omega \sqrt{2(p+1) \log n}\}.$$
Consider the event
\[ \mathcal{A}_0 = \bigcup_{\mathcal{A}_1, \mathcal{A}_2 \in \mathcal{A}_0} \left\{ 0 < \left| \mathbf{X}^\top (\theta_{0, \mathcal{A}_1} - \theta_{0, \mathcal{A}_2}) \right| \leq \frac{4 \sqrt{2(p + 1)c_3 \omega \log n}}{\sqrt{n} \delta_{\min}} \right\}. \]

By Condition (A5) and Bonferroni's inequality, we have
\[ \Pr(\mathcal{A}_0) \leq \sum_{\mathcal{A}_1, \mathcal{A}_2 \in \mathcal{A}_0} \Pr \left( 0 < \left| \mathbf{X}^\top (\theta_{0, \mathcal{A}_1} - \theta_{0, \mathcal{A}_2}) \right| \leq \frac{4 \sqrt{2(p + 1)c_3 \omega \log n}}{\sqrt{n} \delta_{\min}} \right) \]
\[ \leq K^2 \left( \frac{4 \sqrt{2(p + 1)c_3 \omega \log n}}{\sqrt{n} \delta_{\min}} \right)^\gamma. \]  

By the definition of \( \bar{\pi}(\cdot) \), we have
\[ \bar{\pi}(\mathbf{d}) = \mathbb{E} \left( \int_{\mathbf{d}(X)} \mathbf{X}^\top \theta_0(a) \pi^*(a; X, \mathbf{d}(X)) d\mathbf{a} \right). \]

Notice that the expectation in the above expression is taken with respect to \( X \). Define an interval-valued function \( \bar{d}_0(x) = [\tau_{0, \mathbb{R}(x)} - 1, \tau_{0, \mathbb{R}(x)}] \) and set \( \bar{\Delta}(x) = \tilde{d}(x) \cap \{ \bar{d}_0(x) \}^c \). It follows that
\[ \bar{\pi}(\mathbf{d}) = \mathbb{E} \left( \int_{\bar{d}_0(X) \cap \tilde{d}(X)} \mathbf{X}^\top \theta_0(a) \pi^*(a; X, \mathbf{d}(X)) d\mathbf{a} \right) + \mathbb{E} \left( \int_{\bar{\Delta}(X)} \mathbf{X}^\top \theta_0(a) \pi^*(a; X, \mathbf{d}(X)) d\mathbf{a} \right) \]
\[ = \mathbb{E} \left( \int_{\bar{d}_0(X) \cap \tilde{d}(X)} \mathbf{X}^\top \theta_0(a) \pi^*(a; X, \mathbf{d}(X)) d\mathbf{a} \right) + \chi_1. \]

Here, the second equality is due to that \( \pi^*(a; X, \mathbf{d}(X)) = 0 \), for any \( a \in \{ \tilde{d}(X) \}^c \). By (A.12) and (A.65), we have
\[ |\chi_1| \leq c_0 \bar{c}_4 \delta_{\min}^{-1} \mathbb{E} \left( \int_{\bar{\Delta}(X)} ||\mathbf{X}||_2 d\mathbf{a} \right) = c_0 \bar{c}_4 \delta_{\min}^{-1} \mathbb{E} ||\mathbf{X}||_2 \lambda(\bar{\Delta}(X)), \]
where \( \lambda(\bar{\Delta}(X)) \) denotes the Lebesgue measure of \( \bar{\Delta}(X) \). Under the event defined in (A.63), we have \( \lambda(\bar{\Delta}(X)) \leq 2c n^{-1} \log n \), for any realization of \( X \). It follows that
\[ |\chi_1| \leq 2c c_0 \bar{c}_4 \delta_{\min}^{-1} (n^{-1} \log n) \mathbb{E} ||\mathbf{X}||_2. \]  

(A.76)
By (A.33), we have
\[ E\|\mathbf{X}\|_2^2 = \sum_{j=1}^{p+1} E|\mathbf{X}^{(j)}|^2 = \sum_{j=1}^{p+1} E(E|\mathbf{X}^{(j)}|^2|A) \leq \omega^2(p+1). \tag{A.77} \]

By Cauchy-Schwarz inequality, this further implies that
\[ E\|\mathbf{X}\|_2 \leq \sqrt{E\|\mathbf{X}\|_2^2} \leq \omega \sqrt{p + 1}. \]

This together with (A.76) yields
\[ |\chi_1| \leq 2c_0 \tilde{c} \omega \sqrt{p + 1} \delta_{\min}^{-1} n^{-1} \log n, \tag{A.78} \]

with probability at least \(1 - O(n^{-2})\).

Notice that \(\theta_0(\cdot)\) is a constant on \(\hat{d}_0(x)\) for any \(x\). It follows that
\[
\begin{align*}
E\left(\int_{\hat{d}_0(X)} \mathbf{X}^T \theta_0(a) \pi^*(a; X, \hat{d}(X)) da \right) &= E\left(\mathbf{X}^T \theta_0(\tau_{0,R(x)}-1, \tau_{0,R(x)}) \right) \int_{\hat{d}_0(X)} \pi^*(a; X, \hat{d}(X)) da \\
&= E\left(\mathbf{X}^T \theta_0(\tau_{0,R(x)}-1, \tau_{0,R(x)}) \right) \int_{\hat{d}_0(X) \cap \tilde{d}(X)} \pi^*(a; X, \hat{d}(X)) da \\
&= E\left(\mathbf{X}^T \theta_0(\tau_{0,R(x)}-1, \tau_{0,R(x)}) \right) \int_{\tilde{d}(X)} \pi^*(a; X, \hat{d}(X)) da - \chi_2 = E\left(\mathbf{X}^T \theta_0(\tau_{0,R(x)}-1, \tau_{0,R(x)}) \right) - \chi_2,
\end{align*}
\]

where
\[ \chi_2 = E\left(\mathbf{X}^T \theta_0(\tau_{0,R(x)}-1, \tau_{0,R(x)}) \right) \int_{\tilde{d}(X)} \pi^*(a; X, \hat{d}(X)) da. \]

Similar to (A.78), we can show that
\[ |\chi_2| = O(n^{-1} \log n), \]

with probability at least \(1 - O(n^{-2})\). This together with (A.78) yields (A.69).

**Proof of Part 3:** Similar to the definition of \(\mathbb{K}\), we define
\[ \mathbb{K}_0(x) = \text{sarg max}_{k \in \{1, ..., K\}} x^T \theta_0(\tau_{0,k-1}, \tau_{0,k}). \tag{A.79} \]
Let

\[ K^*(x) = \left\{ k_0 : k_0 = \arg \max_{k \in \{1, \ldots, K\}} \bar{x}^\top \theta_{0,[\tau_{0,k-1}, \tau_{0,k})} \right\}, \]

denote the set that consists of all the maximizers. Apparently, \( K_0(x) \in K^*(x) \), \( \forall x \in X \).

We now claim that

\[ \mathbb{R}(X) \in K^*(X), \quad (A.80) \]

under the events defined in \( \mathcal{A}^c_0 \cap \mathcal{A}^* \) and (A.67). Otherwise, suppose there exists some \( k_0 \in \{1, \ldots, K\} \) such that

\[ \bar{X}^\top \hat{\theta}_{1,[\tau_{k_0-1}, \tau_{k_0})} \geq \max_{k \neq k_0} \bar{X}^\top \hat{\theta}_{1,[\tau_{k-1}, \tau_k)}, \quad (A.81) \]

\[ \max_{k \neq k_0} \bar{X}^\top \theta_{0,[\tau_{0,k-1}, \tau_{0,k})} > \bar{X}^\top \theta_{0,[\tau_{0,k_0-1}, \tau_{0,k_0})}. \quad (A.82) \]

Under \( \mathcal{A}^c_0 \), it follows from (A.82) that

\[ \max_{k \neq k_0} \bar{X}^\top \theta_{0,[\tau_{0,k-1}, \tau_{0,k})} > \bar{X}^\top \theta_{0,[\tau_{0,k_0-1}, \tau_{0,k_0})} + \frac{4\sqrt{2(p+1)c_3 \omega \log n}}{\sqrt{n \delta_{\min}}}. \quad (A.83) \]

Under the events defined in \( \mathcal{A}^* \) and (A.67), we have

\[ \max_{k \in \{1, \ldots, K\}} \| \bar{X}^\top (\hat{\theta}_{1,[\tau_{k-1}, \tau_k)} - \theta_{0,[\tau_{0,k-1}, \tau_{0,k})}) \|_2 \leq \| \bar{X} \|_2 \max_{k \in \{1, \ldots, K\}} \| \hat{\theta}_{1,[\tau_{k-1}, \tau_k)} - \theta_{0,[\tau_{0,k-1}, \tau_{0,k})} \|_2 \leq \frac{2\sqrt{2(p+1)c_3 \omega \log n}}{\sqrt{n \delta_{\min}}} \]

This together with (A.83) yields that

\[ \max_{k \neq k_0} \bar{X}^\top \hat{\theta}_{1,[\tau_{k-1}, \tau_k)} > \bar{X}^\top \hat{\theta}_{1,[\tau_{k_0-1}, \tau_{k_0})}. \]

In view of (A.81), we have reached a contradiction. Therefore, (A.80) holds under the events defined in \( \mathcal{A}^c_0 \cap \mathcal{A}^* \) and (A.67). When (A.80) holds, it follows from the definition of \( K^*(\cdot) \) that

\[ \bar{X}^\top \theta_{0,[\mathbb{R}(X)-1, \mathbb{R}(X)]} = \bar{X}^\top \theta_{0,[\mathbb{R}_0(X)-1, \mathbb{R}_0(X)]}. \]

Therefore, under the event defined in (A.67),
we have
\[
E \left( \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) = E \left( \bar{X}^T \theta_0 | \tau_0, R(X), 1 - \tau_0, R(X) \right) I(\mathcal{A}_0^c \cap \mathcal{A}_*)
\] (A.84)
\[
+ E \left( \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) I(\mathcal{A}_0^c \cup \mathcal{A}_*)
\]
\[
+ \chi_3 = E \left( \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) \] (A.84)
\[
+ \chi_4 = E \left( \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) I(\mathcal{A}_0 \cup \mathcal{A}_*)
\].

Notice that
\[
\chi_3 - \chi_4 = E \bar{X}^T \left( \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) - E \left( \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) I(\mathcal{A}_0 \cup \mathcal{A}_*)
\]

Using similar arguments in showing (A.80), we can show that under the event defined in (A.67),
\[
\bar{X}^T \left( \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) - \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) \neq 0,
\]
only when
\[
0 < \left| \bar{X}^T \left( \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) - \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) \right| \leq \frac{4 \sqrt{2(p + 1) c_3 \omega \log n}}{\sqrt{n \delta_{\min}}}
\]

Therefore, under the event defined in (A.67), we have
\[
|\chi_3 - \chi_4| \leq \frac{4 \sqrt{2(p + 1) c_3 \omega \log n}}{\sqrt{n \delta_{\min}}} \Pr(\mathcal{A}_0 \cup \mathcal{A}_*)
\]

It follows from (A.74) and (A.75), we have
\[
|\chi_3 - \chi_4| \leq \frac{4 \sqrt{2(p + 1) c_3 \omega \log n}}{\sqrt{n \delta_{\min}}} \left\{ \frac{2(p + 1)}{n^2} + K^2 \left( \frac{4 \sqrt{2(p + 1) c_3 \omega \log n}}{\sqrt{n \delta_{\min}}} \right)^\gamma \right\}
\]

For sufficiently large \( n \), this together with (A.69) and (A.84) implies that we have with probability at least \( 1 - O(n^{-2}) \),
\[
V^{\tau_*} (\tilde{d}) \geq E \bar{X}^T \theta_0 | \tau_0, \tilde{R}(X), 1 - \tau_0, \tilde{R}(X) \right) - O(1)(n^{-1} \log n + n^{-(1+\gamma)/2} \log^{1+\gamma} n),
\]
for some positive constant $O(1)$. The proof is hence completed by noting that

$$V^{opt} = E\left(\hat{X}^\top \theta_0(\tau_0(X_0))\right).$$

### A.2.7 Proof of Theorem 3

We first introduce some technical lemmas. We remark that the key ingredient of the proof lies in Lemma 5, which establishes a uniform upper bound on the mean squared error of $q_{\mathcal{S},0}$. See Section A.2.8 for a detailed proof. The rest of the proof can be similarly proven as Theorem 1. Specifically, we first show the consistency of the estimated change point locations. We then derive the rate of convergence of the estimated change point locations and the estimated outcome regression function.

**Lemma 5** Assume conditions in Theorem 3 are satisfied. Then there exists some constant $\tilde{C} > 0$ such that the following holds with probability at least $1 - O(n^{-2})$: For any $\mathcal{S} \in \mathcal{I}(m)$ and $|\mathcal{S}| \geq c\gamma_n$,

$$E|q_{\mathcal{S},0}(X) - \tilde{q}_{\mathcal{S}}(X)|^2 \leq \tilde{C}(n|\mathcal{S}|)^{-2\beta/(2\beta+p)}\log^8 n,$$  \hspace{1cm} (A.85)

where $q_{\mathcal{S},0} = E(Y|A \in \mathcal{S}, X)$ for any interval $\mathcal{S}$.

**Lemma 6** Assume conditions in Theorem 3 are satisfied. Then there exists some constant $\tilde{C} > 0$ such that the followings hold with probability at least $1 - O(n^{-2})$: For any $\mathcal{S} \in \mathcal{I}(m)$ and $|\mathcal{S}| \geq c\gamma_n$,

$$\sum_{\mathcal{S} \in \tilde{\mathcal{S}}} \sum_{i=1}^n \mathbb{1}(A_i \in \mathcal{S})|Y_i - q_{\mathcal{S},0}(X_i)||\tilde{q}_{\mathcal{S}}(X_i) - q_{\mathcal{S},0}(X_i)| \leq \tilde{C}(n|\mathcal{S}|)^{p/(2\beta+p)}\log^8 n,$$

for any $\mathcal{S} \in \mathcal{I}(m)$ such that $|\mathcal{S}| \geq c\gamma_n$ for any positive constant $c > 0$.

**Lemma 7** Under the conditions in Theorem 3, the following events occur with probability at least $1 - O(n^{-2})$: there exists some constant $C > 0$ such that $\min_{\mathcal{S} \in \tilde{\mathcal{S}}} |\mathcal{S}| \geq C\gamma_n$.

We next show the consistency of the estimated change-point locations. Using similar arguments in proving (A.16), we can show that

$$|\tilde{\mathcal{S}}| \leq C_0\gamma_n^{-1},$$  \hspace{1cm} (A.86)

for sufficiently large $n$ and some constant $C_0 > 0$. 

148
Notice that
\[
\sum \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[Y_i - \tilde{q}_{\mathcal{S}}(X_i)]^2 \geq \sum \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[Y_i - q_{\mathcal{S}_0}(X_i)]^2 \eta_i
\]
\[+ \sum \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[\tilde{q}_{\mathcal{S}}(X_i) - q_{\mathcal{S}_0}(X_i)]^2
\]
\[-2 \sum \mathbb{I}(A_i \in \mathcal{S})[Y_i - q_{\mathcal{S}_0}(X_i)] \{\tilde{q}_{\mathcal{S}}(X_i) - q_{\mathcal{S}_0}(X_i)\}.
\]

The second line is non-negative. Under Lemmas 6 and 7, the third line is lower bounded by \(-C_1 \sum_{\mathcal{S} \in \mathcal{F}} (n|\mathcal{S}|)^{p/(2\beta+p)} \log^8 n\) for some constant \(C_1 > 0\). By Hölder's inequality, it can be further lower bounded by \(-C_1 \mathbb{P}[2\beta/(2\beta+p) n^{p/(2\beta+p)}] \log^8 n\). By (A.86) and the given condition on \(\gamma_n\), the third line is \(o(n)\). It follows that
\[
\sum \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[Y_i - \tilde{q}_{\mathcal{S}}(X_i)]^2 \geq \eta_i^* + o(n), \tag{A.87}
\]
with probability at least \(1 - O(n^{-2})\).

Similar to (A.9) and (A.10), we can show that the following events occur with probability at least \(1 - O(n^{-2})\),
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[Y_i - Q(X_i, A_i)] \{Q(X_i, A_i) - q_{\mathcal{S}_0}(X_i)\} \right|
\leq c_0 \left[ n^{-1/2} \sqrt{\mathbb{E}[A \in \mathcal{S}][Q(X, A) - q_{\mathcal{S}_0}(X)]^2 \log n + n^{-1} \log n} \right],
\]
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{S})[Q(X_i, A_i) - q_{\mathcal{S}_0}(X_i)]^2 - \mathbb{E}[A \in \mathcal{S}][Q(X, A) - q_{\mathcal{S}_0}(X)]^2 \right|
\leq c_0 \left[ n^{-1/2} \sqrt{\mathbb{E}[A \in \mathcal{S}][Q(X, A) - q_{\mathcal{S}_0}(X)]^2 \log n + n^{-1} \log n} \right],
\]
for some constant \(c_0 > 0\) and any \(\mathcal{S}\). The two upper bounds are \(o(1)\). Similar to (A.20), we can show that
\[
\eta_i^* = \sum_{i=1}^{n} [Y_i - Q(X_i, A_i)]^2 + n \sum_{\mathcal{S} \in \mathcal{F}} \mathbb{E}[A \in \mathcal{S}][Q(X, A) - q_{\mathcal{S}_0}(X)]^2 + o(n),
\]
which completes the proof.
with probability at least $1 - O(n^{-2})$. It follows from (A.87) that
\[
\sum_{\mathcal{F} \in \mathcal{F}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [Y_i - \tilde{q}_\mathcal{F}(X_i)]^2 \geq \sum_{i=1}^{n} [Y_i - Q(X_i, A_i)]^2 + \eta_2^* + n \sum_{\mathcal{F} \in \mathcal{F}} \mathbb{E}(A \in \mathcal{F}) |Q(X, A) - q_{\mathcal{F},0}(X)|^2 + o(n),
\]
(A.88)

with probability at least $1 - O(n^{-2})$.

Let us consider $\eta_2^*$. We observe that
\[
\eta_2^* = \sum_{\mathcal{F} \in \mathcal{F}_0} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [Y_i - q_{\mathcal{F},0}(X_i)]^2.
\]

By the uniform approximation property of DNN, there exists some $q^*_\mathcal{F} \in \mathcal{Q}_\mathcal{F}$ such that
\[
\sum_{i=1}^{n} |q_{\mathcal{F},0}(X_i) - q^*_\mathcal{F}(X_i)|^2 \approx n(n|\mathcal{F}|)^{-2\beta/(2\beta + p)}.
\]

See Part 1 of the proof of Lemma 5 for details. Similar to (A.9) and (A.10), we can show that

the following events occur with probability at least $1 - O(n^{-2})$,
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) \{Y_i - q_{\mathcal{F},0}(X_i)\} \{q_{\mathcal{F},0}(X_i) - q^*_\mathcal{F}(X_i)\} \right| \leq \frac{c_0 \sqrt{|\mathcal{F}| \log n}}{\sqrt{n}} \frac{(n|\mathcal{F}|)^{-\beta/(2\beta + p)}},
\]

for some constant $c_0 > 0$ and any $\mathcal{F} \in \mathcal{F}_0$. It follows that
\[
\eta_2^* - \sum_{\mathcal{F} \in \mathcal{F}_0} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [Y_i - q^*_\mathcal{F}(X_i)]^2 \geq - \sum_{\mathcal{F} \in \mathcal{F}_0} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [q_{\mathcal{F},0}(X_i) - q^*_\mathcal{F}(X_i)]^2
\]
\[-2 \sum_{\mathcal{F} \in \mathcal{F}_0} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) \{Y_i - q_{\mathcal{F},0}(X_i)\} \{q_{\mathcal{F},0}(X_i) - q^*_\mathcal{F}(X_i)\} \geq - \tilde{c} n^{p/(2\beta + p)},
\]

for some constant $\tilde{c} > 0$. This together with (A.88) yields that
\[
\sum_{\mathcal{F} \in \mathcal{F}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [Y_i - \tilde{q}_\mathcal{F}(X_i)]^2 \geq \sum_{\mathcal{F} \in \mathcal{F}_0} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{F}) [Y_i - q^*_\mathcal{F}(X_i)]^2
\]
\[+ n \sum_{\mathcal{F} \in \mathcal{F}} \mathbb{E}(A \in \mathcal{F}) |Q(X, A) - q_{\mathcal{F},0}(X)|^2 + o(n) + O(n^{p/(2\beta + p)})�,
\]

150
with probability at least $1 - O(n^{-2})$.

Next, using similar arguments in proving (A.24), we can show that there exist a partition $\mathcal{P}^* \in \mathcal{B}(m)$ and a set of functions $\{q_{\gamma}^{**} : \gamma \in \mathcal{P}^*\}$ with $|\mathcal{P}^*| = |\mathcal{P}_0|$ such that

$$\sum_{\gamma \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - q_{\gamma}^{**}(X_i)|^2 \geq \sum_{\gamma \in \mathcal{P}^*} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - q_{\gamma}^{**}(X_i)|^2 + O(1).$$

It follows that

$$\sum_{\gamma \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - \tilde{q}_{\gamma}(X_i)|^2 \geq \sum_{\gamma \in \mathcal{P}^*} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - q_{\gamma}^{**}(X_i)|^2 + n \sum_{\gamma \in \mathcal{P}} \mathbb{E}(A \in \gamma) |Q(X, A) - q_{\gamma,0}(X)|^2 + o(n) + O(n^{p/(2\beta+p)}),$$

with probability at least $1 - O(n^{-2})$. Since

$$\sum_{\gamma \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - \tilde{q}_{\gamma}(X_i)|^2 + n \gamma_n |\mathcal{T}| \leq \sum_{\gamma \in \mathcal{P}^*} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - q_{\gamma}^{**}(X_i)|^2 + n \gamma_n |\mathcal{P}_0|,$$

and that $\gamma_n \to 0$, we obtain that

$$\sum_{\gamma \in \mathcal{P}} \mathbb{E}(A \in \gamma) |Q(X, A) - q_{\gamma,0}(X)|^2 = o(1).$$

Under the condition that $q_{\gamma,1} \neq q_{\gamma,0}$ for any adjacent $\gamma_1, \gamma_2 \in \mathcal{P}_0$, we have $\mathbb{E}|q_{\gamma,0}(X) - q_{\gamma,0}(X)|^2 > 0$. Using similar arguments in the Part 1 of the proof of Theorem 1, we obtain that $\max_{\gamma \in \mathcal{P}} \min_{\tau \in \mathcal{P}} |\tau - \gamma| \leq \delta$ for any constant $\delta > 0$. This further implies that $|\mathcal{T}| \geq |\mathcal{P}_0|.

We next derive the rate of convergence of the estimated change point locations and the estimated outcome regression function. Similar to (A.89), with a more refined analysis (see e.g., Step 2 of the proof of Theorem 1), we obtain that

$$\sum_{\gamma \in \mathcal{P}} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - \tilde{q}_{\gamma}(X_i)|^2 \geq \sum_{\gamma \in \mathcal{P}^*} \sum_{i=1}^{n} \mathbb{I}(A_i \in \gamma) |Y_i - q_{\gamma}^{**}(X_i)|^2 + n \sum_{\gamma \in \mathcal{P}} \mathbb{E}(A \in \gamma) |Q(X, A) - q_{\gamma,0}(X)|^2 - C_1 |\mathcal{T}|^{2\beta/(2\beta+p)} n^{p/(2\beta+p)} \log^8 n + O(n^{p/(2\beta+p)}),$$

151
with probability at least $1 - O(n^{-2})$. This together with (A.90) yields that
\[
 n \sum_{\mathcal{I} \in \mathcal{P}} \mathbb{E}\{A \in \mathcal{I}|Q(X, A) - q_{\mathcal{I}, 0}(X)\}^2 \leq C_1 \left| \hat{\mathcal{D}} \right|^{2\beta/(2\beta+p)} n^{p/(2\beta+p)} \log^8 n \\
+ O(n^{p/(2\beta+p)}) + n\gamma n(|\mathcal{P}_0| - |\hat{\mathcal{D}}|).
\]
Under the given condition on $\gamma_n$, we obtain that $|\hat{\mathcal{D}}| \leq |\mathcal{P}_0|$. Combining this together with $|\mathcal{D}| \geq |\mathcal{P}_0|$, we obtain that $|\mathcal{D}| = |\mathcal{P}_0|$. This proves the results in (i).

Consequently, we obtain that
\[
 n \sum_{\mathcal{I} \in \mathcal{P}} \mathbb{E}\{A \in \mathcal{I}|Q(X, A) - q_{\mathcal{I}, 0}(X)\}^2 = O(n^{p/(2\beta+p)} \log^8 n),
\]
As such, we have that
\[
 \sum_{\mathcal{I} \in \mathcal{P}} \mathbb{E}\{A \in \mathcal{I}|Q(X, A) - q_{\mathcal{I}, 0}(X)\}^2 = O(n^{-2\beta/(2\beta+p)} \log^8 n),
\]
This together with Lemma 5 proves the result in (iii). Using similar arguments in Part 2 of the proof of Theorem 1, we can show the result in (ii) holds. This completes the proof.

### A.2.8 Proof of Lemma 5

By definition, $\hat{q}_\mathcal{I}$ is the minimizer of the least square loss, argmin$_{q \in \mathcal{Q}_\mathcal{I}} \sum_{i=1}^n I(A_i \in \mathcal{I})|Y_i - q(X_i)|^2$. It follows that
\[
 \sum_{i=1}^n I(A_i \in \mathcal{I})|Y_i - \hat{q}_\mathcal{I}(X_i)|^2 \geq \sum_{i=1}^n I(A_i \in \mathcal{I})|Y_i - q(X_i)|^2,
\]
for all $q \in \mathcal{Q}_\mathcal{I}$. Let $q_{\mathcal{I}, 0}(x) = \mathbb{E}[Y|A \in \mathcal{I}, X = x]$, we have $\mathbb{E}[I(A \in \mathcal{I})|Y - q_{\mathcal{I}, 0}(X)|X] = 0$. A simple calculation yields
\[
 \sum_{i=1}^n I(A_i \in \mathcal{I})|q_{\mathcal{I}, 0}(X_i) - \hat{q}_\mathcal{I}(X_i)|^2 \leq \sum_{i=1}^n I(A_i \in \mathcal{I})|q_{\mathcal{I}, 0}(X_i) - q(X_i)|^2 \\
+ 2 \sum_{i=1}^n I(A_i \in \mathcal{I})\{Y_i - q_{\mathcal{I}, 0}(X_i)\}\{\hat{q}_\mathcal{I}(X_i) - q_{\mathcal{I}, 0}(X_i)\},
\]
for any $q$ and $\mathcal{I}$.

The first term on the RHS measures the approximation bias of the DNN class. Since
E[\mathbb{I}(A \in \mathcal{G})|Y - q_{\mathcal{S},0}(X)|X] = 0, the second term corresponds to the stochastic error. The rest of the proof is divided into three parts. In Part 1, we bound the approximation error. In Part 2, we bound the stochastic error. Finally, we combine these two parts together to derive the uniform convergence rate for \( \hat{q}_\mathcal{S} \).

**Part 1.** Under the given condition, we have \( Q(\bullet, a) \in \Phi(\beta, c) \), \( p(a|\bullet) \in \Phi(\beta, c) \) for some \( c > 0 \) and any \( a \). We now argue that there exists some constant \( C > 0 \) such that \( q_{\mathcal{S}} \in \Phi(\beta, C) \) for any \( \mathcal{S} \). This can be proven based on the relation that

\[
q_{\mathcal{S}}(x) = \frac{\int_{\mathcal{S}} Q(x, a)p(a|x)da}{\int_{\mathcal{S}} p(a|x)da}.
\]

Specifically, we have that \( \sup_{x} |q_{\mathcal{S}}(x)| \leq \sup_{a,x} |Q(x, a)| \leq c \). Suppose \( \beta \leq 1 \). For any \( x_1, x_2 \in X \), consider the difference \( |q_{\mathcal{S}}(x_1) - q_{\mathcal{S}}(x_2)| \). Under the assumption that \( \inf_{a,x} p(a|x) \geq c_a \), it follows that

\[
|q_{\mathcal{S}}(x_1) - q_{\mathcal{S}}(x_2)| \leq \frac{\int_{\mathcal{S}} |Q(x_1, a) - Q(x_2, a)|p(a|x_1)da}{\int_{\mathcal{S}} p(a|x_1)da} + \frac{\int_{\mathcal{S}} |Q(x_2, a)||p(a|x_1) - p(a|x_2)|da}{\int_{\mathcal{S}} p(a|x_1)da} + \frac{\int_{\mathcal{S}} |Q(x_1, a)||p(a|x_2) - p(a|x_1)|da}{\int_{\mathcal{S}} p(a|x_2)da} + \frac{\int_{\mathcal{S}} |Q(x_2, a)||p(a|x_1) - p(a|x_2)|da}{\int_{\mathcal{S}} p(a|x_2)da} \leq c \|x_1 - x_2\|^{\beta - 1} + \frac{2c^2}{c_a} \|x_1 - x_2\|^{\beta - 1}.
\]

Consequently, \( q_{\mathcal{S}} \in \Phi(\beta, c + 2c^2/c_a^2) \).

Suppose \( \beta > 1 \). Then both \( Q(\bullet, a) \) and \( p(a|\bullet) \) are \( |\beta| \)-differentiable. By changing the order of integration and differentiation, we can show that \( q_{\mathcal{S}}(x) \) is \( |\beta| \)-differentiable as well. As an illustration, when \( \beta < 2 \), we have \( |\beta| = 1 \). According to the chain rule, we have

\[
\frac{\partial q_{\mathcal{S}}(x)}{\partial x^i} = \frac{\int_{\mathcal{S}} \{\partial Q(x, a)/\partial x^i\}p(a|x)da}{\int_{\mathcal{S}} p(a|x)da} + \frac{\int_{\mathcal{S}} Q(a|x)\{\partial p(a|x)/\partial x^i\}da}{\int_{\mathcal{S}} p(a|x)da} - \frac{\int_{\mathcal{S}} Q(a|x)p(a|x)da}{\int_{\mathcal{S}} p(a|x)da} \frac{\int_{\mathcal{S}} \{\partial p(a|x)/\partial x^i\}da}{\int_{\mathcal{S}} p(a|x)da}.
\]

Moreover, using similar arguments in proving \( q_{\mathcal{S}} \in \Phi(\beta, c + 2c^2/c_a^2) \) when \( \beta < 1 \), we can show that all the partial derivatives of \( q_{\mathcal{S}}(x) \) up to the \( |\beta| \)th order are uniformly bounded for all \( \mathcal{S} \). In addition, all the \( |\beta| \)th order partial derivatives are Hölder continuous with exponent \( \beta - |\beta| \). This implies that \( q_{\mathcal{S}} \in \Phi(\beta, C) \) for some constant \( C > 0 \) and any \( \mathcal{S} \).
It is shown in Lemma 7 of Farrell et al. (2021) that for any \( \epsilon > 0 \), there exists a DNN architecture that approximates \( q_\mathcal{F} \) with the uniform approximation error upper bounded by \( \epsilon \), and satisfies \( W_\mathcal{F} \leq \tilde{C} e^{-p/\beta}(\log \epsilon^{-1}+1) \) and \( L_\mathcal{F} \leq \tilde{C}(\log \epsilon^{-1}+1) \) for some constant \( \tilde{C} > 0 \). These upper bounds will be used later in Part 2. The detailed value of \( \epsilon \) will be specified below. It follows that for any \( \mathcal{F} \), the bias term can be upper bounded by

\[
\sum_{i=1}^{n} \mathbb{I}(A_{i} \in \mathcal{F})|q_{\mathcal{F},0}(X_{i}) - q(X_{i})|^{2} \leq \epsilon^{2} \sum_{i=1}^{n} \mathbb{I}(A_{i} \in \mathcal{F}).
\]

Using similar arguments in proving (A.10), we can show that uniformly for all \( \mathcal{F} \) such that \( |\mathcal{F}| \geq c \gamma_{n} \), the RHS can be upper bounded by \( O(1)(n|\mathcal{F}| + \sqrt{n|\mathcal{F}| \log n}) \) with probability tending to 1, where \( O(1) \) denotes some universally constant that is independent of \( \mathcal{F} \). It follows that uniformly for all \( \mathcal{F} \) such that \( |\mathcal{F}| \geq c \gamma_{n} \), there exists a DNN with \( W_\mathcal{F} \leq \tilde{C} e^{-p/\beta}(\log \epsilon^{-1}+1) \) and \( L_\mathcal{F} \leq \tilde{C}(\log \epsilon^{-1}+1) \) such that

\[
\sum_{i=1}^{n} \mathbb{I}(A_{i} \in \mathcal{F})|q_{\mathcal{F},0}(X_{i}) - q(X_{i})|^{2} \propto \epsilon^{2}(n|\mathcal{F}| + \sqrt{n|\mathcal{F}| \log n}),
\]

for any sufficiently small \( \epsilon > 0 \).

Set \( \epsilon \) to \( (n|\mathcal{F}|)^{-\beta/(2\beta+p)} \), it follows that

\[
\sum_{i=1}^{n} \mathbb{I}(A_{i} \in \mathcal{F})|q_{\mathcal{F},0}(X_{i}) - q(X_{i})|^{2} \propto (n|\mathcal{F}|)^{-2\beta/(2\beta+p)}(n|\mathcal{F}| + \sqrt{n|\mathcal{F}| \log n}). \quad (A.91)
\]

\( W_\mathcal{F} \) and \( L_\mathcal{F} \) are upper bounded by \( \tilde{C}(n|\mathcal{F}|)^{p/(2\beta+p)}(\beta \log(n|\mathcal{F}|)/(2\beta+p)+1) \) and \( \tilde{C}(\beta \log(n|\mathcal{F}|)/(2\beta+p)+1) \), respectively. This completes the proof for Part 1.

**Part 2.** We will apply the empirical process theory (see e.g., Van Der Vaart and Wellner 1996) to bound the stochastic error. Let \( \hat{\theta}_{\mathcal{F}} \) be the estimated parameter in \( \hat{q}_{\mathcal{F}} \). Define

\[
\sigma^{2}(\mathcal{F}, \theta) = \mathbb{E}(A \in \mathcal{F})|q_{\mathcal{F},0}(X) - q_{\mathcal{F}}(X, \theta)|^{2},
\]

for any \( \theta \) and \( \mathcal{F} \).

Consider two separate cases, corresponding to \( \sigma(\mathcal{F}, \hat{\theta}_{\mathcal{F}}) \leq k_{0}|\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}\log^{4} n \) and \( \sigma(\mathcal{F}, \hat{\theta}_{\mathcal{F}}) > k_{0}|\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}\log^{4} n \), respectively, for some constant \( k_{0} > 0 \). The detailed form of \( k_{0} \) will be specified later in Part 3. We focus our attentions on the latter...
class of intervals. In Part 3, we will show that for those intervals, with proper choice of \( k_0 \),
\[
\Pr(\sigma(\mathcal{G}, \hat{\theta}_\mathcal{G}) \leq k_0|\mathcal{G}|^{1/2}(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4 n) \geq 1 - O(n^{-4}).
\]
By Bonferroni’s inequality and the condition that \( m \approx n \), this implies that for any \( \mathcal{G} \), we have
\[
\sigma(\mathcal{G}, \hat{\theta}_\mathcal{G}) \leq k_0|\mathcal{G}|^{1/2}(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4 n,
\]
with probability at least \( 1 - O(n^{-2}) \).

For a given integer \( k \geq k_0 \). We consider the stochastic error,
\[
\sum_{i \in \mathcal{L}_t^\mathcal{G}} \mathbb{I}(A_i \in \mathcal{G})\{Y_i - q_{\mathcal{G}, 0}(X_i)\}\{q_{\mathcal{G}, 0}(X_i) - q_{\mathcal{G}, 0}(X_i)\},
\]
for any \( \mathcal{G} \) such that
\[
k|\mathcal{G}|^{1/2}(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4 n \leq \sigma(\mathcal{G}, \hat{\theta}_\mathcal{G})
\leq (k + 1)|\mathcal{G}|^{1/2}(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4 n.
\]
The absolute value of the stochastic error can be upper bounded by
\[
Z(\mathcal{G}) \equiv \sup_{\theta} \left| \sum_{i \in \mathcal{L}_t^\mathcal{G}} \mathbb{I}(A_i \in \mathcal{G})\{Y_i - q_{\mathcal{G}, 0}(X_i)\}\{q_{\mathcal{G}, 0}(X_i, \theta) - q_{\mathcal{G}, 0}(X_i)\} \right|,
\]
where the supremum is taken over all \( \theta \) such that (A.93) holds with \( \hat{\theta}_\mathcal{G} \) replaced by \( \theta \).

For a given \( \theta \), the empirical sum has zero mean. Under the boundedness assumption on \( Y \), its standard deviation is upper bounded by \( O(1)k|\mathcal{G}|^{1/2}(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4 n \) for some universal constant \( O(1) \). In addition, each quantity \( \mathbb{I}(A_i \in \mathcal{G})\{Y_i - q_{\mathcal{G}, 0}(X_i)\}\{q_{\mathcal{G}, 0}(X_i, \theta) - q_{\mathcal{G}, 0}(X_i)\} \) is upper bounded by some universal constant. This allows us to apply the tail inequality developed by Massart et al. (2000) to bounded the empirical process. See also Theorem 2 of Adamczak et al. (2008). Specifically, for all \( t > 0 \) and \( \mathcal{G} \), we obtain with probability at least \( 1 - \exp(t) \) that
\[
Z(\mathcal{G}) \leq 2 EZ(\mathcal{G}) + \tilde{c} k \sqrt{t n|\mathcal{G}|(n|\mathcal{G}|)^{-\beta/(2\beta+p)}\log^4(n) + \tilde{c} t},
\]
for some constant \( \tilde{c} > 0 \). By setting \( t = 4\log(nk) \), we have \( 1 - \exp(t) = 1 - n^{-4}k^{-4} \). Notice
that the number of intervals $\mathcal{I}$ is upper bounded by $O(n^2)$, under the condition that $m$ is proportional to $n$. By Bonferroni's inequality, we obtain that (A.94) holds with probability at least $1 - O(n^2)$ for any $\mathcal{I}$. Under the given condition on $\gamma$, for any interval $\mathcal{I}$ such that $|\mathcal{I}| \geq c \gamma n$, the last term on the right-hand-side of (A.94) is dominated by the second term. It follows that the following occurs with probability $1 - O(k^{-2} n^{-2})$,

$$Z(\mathcal{I}) \leq 2E_Z(\mathcal{I}) + 3\bar{c} k \sqrt{n|\mathcal{I}|} |(n|\mathcal{I}|)^{-\beta/(2\beta+p)} \sqrt{\log(nk) \log^4 n},$$  \hspace{1cm} (A.95)

for all $\mathcal{I}$ such that $|\mathcal{I}| \geq c \gamma n$ and that (A.93) holds.

We next provide an upper bound for $E_Z(\mathcal{I})$. Toward that end, we will apply the maximal inequality developed in Corollary 5.1 of Chernozhukov et al. (2014). We first observe that the class of empirical sum indexed by $\theta$ belongs to the VC subgraph class with VC-index upper bounded by $O(W_{\mathcal{I}} L_{\mathcal{I}} \log(W_{\mathcal{I}}))$. It follows that for any $\mathcal{I}$ such that $|\mathcal{I}| \geq c \gamma n$ and (A.93) holds,

$$E_Z(\mathcal{I}) \propto \sqrt{k(n|\mathcal{I}|)^{p/(2\beta+p)} W_{\mathcal{I}} L_{\mathcal{I}} \log(W_{\mathcal{I}}) \log(nk) \log^4 n + W_{\mathcal{I}} L_{\mathcal{I}} \log(W_{\mathcal{I}}) \log(nk)}.$$

Based on the upper bounds on $W_{\mathcal{I}}$ and $L_{\mathcal{I}}$ developed in Part 1, the right-hand-side is upper bounded by

$$O(1) \sqrt{k(n|\mathcal{I}|)^{p/(2\beta+p)} \sqrt{\log(nk) \log^{11/2} n},$$

where $O(1)$ denotes some universal constant.

This together with (A.94) and (A.95) yields that with probability at least $1 - O(n^{-2} k^{-2})$, the stochastic error is of the order $\sqrt{k(n|\mathcal{I}|)^{p/(2\beta+p)} \sqrt{\log(nk) \log^{11/2} n}$, for any $\mathcal{I}$ that satisfies $|\mathcal{I}| \geq c \gamma n$ and (A.93). This completes the proof for Part 2.

**Part 3.** Combining the results in Part 1 and Part 2, we obtain that for any $\mathcal{I}$ such that $|\mathcal{I}| \geq c \gamma n$, $k|\mathcal{I}|^{1/2} (n|\mathcal{I}|)^{-\beta/(2\beta+p)} \log^4 n \leq \sigma(\mathcal{I}, \hat{\theta}_{\mathcal{I}}) \leq (k + 1)|\mathcal{I}|^{1/2} (n|\mathcal{I}|)^{-\beta/(2\beta+p)} \log^4 n$,

$$\sum_{i \in L_{\mathcal{I}}} \mathbb{I}(A_i \in \mathcal{I}) q_{\mathcal{I},0}(X_i) - \hat{q}_{\mathcal{I}}^{(i)}(X_i)^2 \leq O(1) \sqrt{k(n|\mathcal{I}|)^{p/(2\beta+p)} \sqrt{\log(nk) \log^{11/2} n},$$
with probability at least $1 - O(n^{-2}k^{-2})$. As for the left-hand-side, we notice that

$$\sum_{i \in L_i^c} \mathbb{I}[A_i \in \mathcal{I}]|q_{\mathcal{I}, 0}(X_i) - \hat{q}_{\mathcal{I}}^{(l)}(X_i)|^2$$

$$\geq |L_i^c| \sigma^2(\mathcal{I}, \hat{\theta}_\mathcal{I}) - \left| \sum_{i \in L_i^c} \mathbb{I}[A_i \in \mathcal{I}]|q_{\mathcal{I}, 0}(X_i) - \hat{q}_{\mathcal{I}}^{(l)}(X_i)|^2 - |L_i^c| \sigma^2(\mathcal{I}, \hat{\theta}_\mathcal{I}) \right|$$

$$= \mathcal{L}^{-1}(\mathcal{L} - 1)k(n|\mathcal{I}|)^{p/(2\beta + p)} \log^8 n \left| \sum_{i \in L_i^c} \mathbb{I}[A_i \in \mathcal{I}]|q_{\mathcal{I}, 0}(X_i) - \hat{q}_{\mathcal{I}}^{(l)}(X_i)|^2 - |L_i^c| \sigma^2(\mathcal{I}, \hat{\theta}_\mathcal{I}) \right| .$$

Using similar arguments in Part 2, we can show that the second term in the last line is of the order $O((\sqrt{k(n|\mathcal{I}|)^{p/(2\beta + p)} \sqrt{\log(nk)}})^{1/2} n)$, with probability at least $1 - O(n^{-2}k^{-2})$, for any $\mathcal{I}$ such that $|\mathcal{I}| \geq c \gamma_n$ and (A.93) holds. Since $\mathcal{L} \geq 2$, we obtain

$$k(n|\mathcal{I}|)^{p/(2\beta + p)} \log^8 n \leq O(1)\sqrt{k(n|\mathcal{I}|)^{p/(2\beta + p)} \sqrt{\log(nk)}}^{1/2} n .$$

For a sufficiently large $k_0$, the above equation will not hold for any $k \geq k_0$. As such, the probability

$$\text{Pr}\left(k|\mathcal{I}|^{1/2}(n|\mathcal{I}|)^{-\beta/(2\beta + p)} \log^4 n \leq \sigma(\mathcal{I}, \hat{\theta}_\mathcal{I}) \leq (k + 1)|\mathcal{I}|^{1/2}(n|\mathcal{I}|)^{-\beta/(2\beta + p)} \log^4 n \right) \leq O(n^{-2}k^{-2}) ,$$

for any $k \geq k_0$. It follows from Bonferroni’s inequality that

$$\text{Pr}\left(\sigma(\mathcal{I}, \hat{\theta}_\mathcal{I}) \geq k_0|\mathcal{I}|^{1/2}(n|\mathcal{I}|)^{-\beta/(2\beta + p)} \log^4 n \right) \leq O \left( \sum_{k = k_0}^{+\infty} n^{-2}k^{-2} \right) = O(k_0n^{-2}) .$$

We thus obtain (A.92) holds with probability at least $1 - O(n^{-2})$. Under the assumption that the density function $b(a|x)$ is uniformly bounded away from zero, we obtain

$$\sigma^2(\mathcal{I}, \hat{\theta}_\mathcal{I}) \leq c \mathcal{I}|E|q_{\mathcal{I}, 0}(X) - \hat{q}_{\mathcal{I}}^{(l)}(X)^2 ,$$

for some constant $c > 0$. This assertion thus follows.

**A.2.9 Proof of Lemma 6**

The assertion can be proven in a similar manner as Part 2 of the proof of Lemma 5. We omit the details to save space.
A.2.10 Proof of Lemma 7

The assertion can be proven in a similar manner as Lemma 4. We omit the details to save space.

A.2.11 Proof of Theorem 4

The proof of Theorem 4 is similar to that of Theorem 2. We provide the outline as below and omit the duplicated arguments for brevity.

Under the events defined in Theorem 3, we have $\hat{K} = K$, and

$$
\max_{k \in \{1, \ldots, K-1\}} |\hat{\tau}_{k} - \tau_{0,k}| \leq c n^{-2\beta/(2\beta + p)} \log^8 n,
$$

(A.96)

for some constant $c > 0$. By similar arguments in the proof of Theorem 2, there exists some constant $\bar{C}_4 > 0$ such that

$$
\pi^*(a; x, \hat{d}(x)) \leq \bar{C}_4 \delta^{-1}, \quad \forall a \in [0, 1], x \in \mathcal{X}.
$$

(A.97)

The rest of our proof is divided into two parts. In the first part, we focus on proving

$$
V^{\pi^*}(\hat{d}) \geq E\left(q_{\tau_{0,\mathcal{X}(x)}-1, \tau_{0,\mathcal{X}(x)}}(X)\right) - O(1)n^{-2\beta/(2\beta + p)} \log^8 n,
$$

(A.98)

with probability at least $1 - O(n^{-2})$, where $O(1)$ denotes some positive constant.

In Part 2, we provide an upper bound for

$$
V^{opt} - E\left(q_{\tau_{0,\mathcal{X}(x)}-1, \tau_{0,\mathcal{X}(x)}}(X)\right).
$$

This together with (A.98) yields the desired results.

Proof of Part 1: Recall the integer-valued function

$$
\hat{K}(x) = \text{sargmax}_{k \in \{1, \ldots, K\}} \hat{q}_{\tau_{k-1}, \tau_k}(x),
$$

(A.99)

where sargmax denotes the smallest maximizer when the argmax is not unique. Similarly, we have $\hat{K}(x) = k$ if $\hat{d}(x) = [\hat{\tau}_{k-1}, \hat{\tau}_k]$ for some integer $k$ such that $1 \leq k \leq K - 1$, and set $\hat{K}(x) = K$ if $\hat{d}(x) = [\hat{\tau}_{K-1}, 1]$. Let $\hat{\Delta}_k = [\hat{\tau}_{k-1}, \hat{\tau}_k] \cup [\tau_{0,k-1}, \tau_{0,k})^c + [\hat{\tau}_{k-1}, \hat{\tau}_k]^c \cup [\tau_{0,k-1}, \tau_{0,k})$. Using similar arguments in
the proof of Theorem 2, we have

\[ V^{\pi^*}(\tilde{d}) = \mathbb{E}\left( \int_{\tilde{d}_0(X) \cap \tilde{d}(X)} Q(X, a)\pi^*(a; X, \tilde{d}(X)) da \right) + \mathbb{E}\left( \int_{\Delta(X)} Q(X, a)\pi^*(a; X, \tilde{d}(X)) da \right) \]

\[ = \mathbb{E}\left( \int_{\tilde{d}_0(X)} Q(X, a)\pi^*(a; X, \tilde{d}(X)) da \right) + \chi_1^*, \]

where \( \tilde{d}_0(x) = [\tau_{0,\mathbb{R}(x) - 1}, \tau_{0,\mathbb{R}(x)}] \) and \( \Delta(x) = \tilde{d}(x) \cap \{ \tilde{d}_0(x) \}^c \).

By (A.97) and the assumption that \( Y \) is bounded, we have

\[ |\chi_1^*| \leq c_0 \tilde{C}_1 \delta_{\min}^{-1} \lambda(\Delta(X)), \]

where \( \lambda(\Delta(X)) \) denotes the Lebesgue measure of \( \Delta(X) \). Under the event defined in (A.96), we have \( \lambda(\Delta(X)) \leq 2n^{-2\beta/(2\beta+p)} \log^8 n \), for any realization of \( X \). It follows that

\[ |\chi_1^*| \leq \tilde{C}_0 \delta_{\min}^{-1} n^{-2\beta/(2\beta+p)} \log^8 n, \quad (A.100) \]

for some constant \( \tilde{C}_0 \) with probability at least \( 1 - O(n^{-2}) \).

Using similar arguments in the proof of Theorem 2, we have

\[ \mathbb{E}\left( \int_{\tilde{d}_0(X)} Q(X, a)\pi^*(a; X, \tilde{d}(X)) da \right) = \mathbb{E}\left( q_{[\tau_{0,\mathbb{R}(x) - 1}, \tau_{0,\mathbb{R}(x)}]}(X) \right) - \chi_2^*, \]

where

\[ \chi_2^* = \mathbb{E}\left( q_{[\tau_{0,\mathbb{R}(x) - 1}, \tau_{0,\mathbb{R}(x)}]}(X) \right) \int_{\Delta(X)} \pi^*(a; X, \tilde{d}(X)) da. \]

Similar to (A.100), we can show that

\[ |\chi_2^*| = O(n^{-2\beta/(2\beta+p)} \log^8 n), \]

with probability at least \( 1 - O(n^{-2}) \). This together with (A.100) yields (A.98).

**Proof of Part 2:** Let \( \epsilon_n = \tilde{C}_1(n \delta_{\min})^{-2\beta/(2\beta+p)(2\gamma)} \log^{8/(2+\gamma)} n \) for some constant \( \tilde{C}_1 \). Define an event

\[ \mathcal{A}_\epsilon = \bigcup_k \left\{ |q_{[\tau_{k-1,\mathbb{R}(x)}]}(X) - \hat{q}_{[\tau_{k-1,\mathbb{R}(x)}]}(X)| \leq \epsilon_n \right\}. \]
Based on Lemma 5, by Markov’s inequality, we can show that there exists some constant \( \tilde{c} > 0 \) such that

\[
\Pr\{|q_{\tau_k, k-1, \tau_k}(X) - \tilde{q}_{\tau_k, k-1, \tau_k}(X)| > \epsilon_n\} \leq \tilde{C}_2(n \delta_{\min})^{-2\beta(1+\gamma)/(2\beta+p)(2+\gamma)} \log^{8(1+\gamma)/(2+\gamma)} n, \forall k \in \{1, \ldots, K\},
\]

with probability at least \( 1 - O(n^{-2}) \) for some constant \( \tilde{C}_2 \). Thus, by Bonferroni’s inequality, we have

\[
\Pr\{A^c\} \leq \tilde{C}_3(n \delta_{\min})^{-2\beta(1+\gamma)/(2\beta+p)(2+\gamma)} \log^{8(1+\gamma)/(2+\gamma)} n
\]

(A.102)

holds with probability at least \( 1 - O(n^{-2}) \) for some constant \( \tilde{C}_3 \).

Consider the event

\[
A_0 = \bigcup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}_0} \{0 < |q_{\mathcal{S}_1, 0}(X) - q_{\mathcal{S}_2, 0}(X)| \leq 2\epsilon_n\}.
\]

By Condition (A5) and Bonferroni’s inequality, we have

\[
\Pr(A_0) \leq \sum_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}_0} \Pr(0 < |q_{\mathcal{S}_1, 0}(X) - q_{\mathcal{S}_2, 0}(X)| \leq 2\epsilon_n) \leq K^2(2\epsilon_n)^\gamma.
\]

(A.103)

Similar to the definition of \( \tilde{K} \), we define

\[
K_0(x) = \text{sarg max}_{k \in \{1, \ldots, K\}} q_{\tau_k, k-1, \tau_k}(x).
\]

(A.104)

Let

\[
K^*(x) = \left\{ k_0 : k_0 = \text{arg max}_{k \in \{1, \ldots, K\}} q_{\tau_k, k-1, \tau_k}(x) \right\},
\]

denote the set that consists of all the maximizers. Apparently, \( K_0(x) \in K^*(x), \forall x \in X \).

We now claim that

\[
\tilde{K}(x) \in K^*(x),
\]

(A.105)

under the events defined in \( A^c_0 \) and \( A^c_\epsilon \). Otherwise, suppose there exists some \( k_0 \in \{1, \ldots, K\} \)
such that

\[
\hat{q}_{[\tau_{k_0}, \tau_{k_0})}(X) \geq \max_{k \neq k_0} \hat{q}_{[\tau_{k-1}, \tau_k)}(X), \tag{A.106}
\]

\[
\max_{k \neq k_0} q_{[\tau_{0,k-1}, \tau_{0,k})} o(0) > q_{[\tau_{0,k_0-1}, \tau_{0,k_0})} o(0). \tag{A.107}
\]

Under \( \mathcal{E}_0 \), it follows from (A.107) that

\[
\max_{k \neq k_0} q_{[\tau_{0,k-1}, \tau_{0,k})} o(0) > q_{[\tau_{0,k_0-1}, \tau_{0,k_0})} o(0) + 2\epsilon_n. \tag{A.108}
\]

Under the event \( \mathcal{E}_e \), we have

\[
\max_{k \in \{1, \ldots, K\}} |\hat{q}_{[\tau_{k-1}, \tau_k)}(X) - q_{[\tau_{0,k-1}, \tau_{0,k})} o(0)| \leq \epsilon_n.
\]

This together with (A.108) yields that

\[
\max_{k \neq k_0} \hat{q}_{[\tau_{k-1}, \tau_k)}(X) > \hat{q}_{[\tau_{k_0-1}, \tau_{k_0})}(X).
\]

In view of (A.106), we have reached a contradiction. Therefore, (A.105) holds under the events defined in \( \mathcal{E}_0 \) and \( \mathcal{E}_e \).

By the definition of \( \mathbb{K}^*(\cdot) \) that \( q_{[\tau_{0,R_1-1}, \tau_{0,R_1})} o(0) = q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) \) when (A.105) holds. Using the similar augments in (A.84), we have

\[
E\left( q_{[\tau_{0,R_1-1}, \tau_{0,R_1})} o(0) \right) = E\left( q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) \right) + \chi_3 + \chi_4, \tag{A.109}
\]

where

\[
\chi_3 = E\left( q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) - q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) \right) \mathbb{I}(\mathcal{E}_0) \mathbb{I}(\mathcal{E}_e),
\]

and

\[
\chi_4 = E\left( q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) - q_{[\tau_{0,R_0-1}, \tau_{0,R_0})} o(0) \right) \mathbb{I}(\mathcal{E}_e).
\]

Therefore, under the event \( \mathcal{E}_e \), it follows from (A.103) that

\[
|\chi_3| \leq K^2 (2\epsilon_n)^{\gamma+1}. \tag{A.110}
\]

Similarly, by Condition (A7) and the outcome is bounded, following Markov's inequality,
we have

\[ |X_4| \leq \hat{C}_3 \Pr\{A^C\} \quad (\text{A.111}) \]

Based on (A.102) and \( \epsilon_n = \hat{C}_1 \left( n \delta_{\min}^{-2\beta/(2\beta+p)(2+\gamma)} \right) \log^{8/(2+\gamma)} n \), for sufficiently large \( n \), the above (A.111) and (A.110) together with (A.98) and (A.109) implies that we have with probability at least \( 1 - O(n^{-2}) \),

\[ V^{\pi^*}(\hat{d}) \geq V^{\alpha_{opt}} - O(1)(n^{-2\beta/(2p+\gamma \ln\beta}) \log^8 n + n^{-2\beta/(2p+\gamma \ln\beta \gamma)} \log^{8+2\gamma} n), \]

for some positive constant \( O(1) \). The proof is hence completed.

**A.2.12 Proof of Theorem 5**

We focus on proving Theorem 5 (ii) when conditions in Theorem 4 are satisfied with \( 4\beta(1 + \gamma) > (2\beta + p)(2 + \gamma) \), where D-JIL is applied. Since the piecewise linear case requires weaker conditions (when conditions in Theorem 2 are satisfied), one can similarly derive the asymptotic normality of \( \hat{\beta} \) under L-JIL.

We present an outline of the proof first, which can be divided into two parts. Define \( d_0(x) = \arg \max_{z \in \mathcal{X}} q_{z,0}(x), \forall x \in \mathcal{X} \). Under the given conditions, the maximizers \( d_0(X) \)'s are almost surely unique. By the definition of \( \hat{K}(\cdot) \) in (A.99), we have

\[ \hat{V} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\mathbb{I}\{A_i \in \hat{d}(X_i)\}}{\hat{e}(\hat{d}(X_i)|X_i)} \{ Y_i - X_i \right] + \hat{q}_{\tau_{0,\hat{K}(X_i)-1,\hat{K}(X_i)}}(X_i) + \hat{q}_{\tau_{0,\hat{K}(X_i)-1,\hat{K}(X_i)}}(X_i) \].

Given \( K_0(\cdot) \) defined in (A.104), the above value estimator can be decomposed by

\[ \hat{V} = \hat{V}_1 + \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\mathbb{I}\{A_i \in \hat{d}(X_i)\}}{\hat{e}(\hat{d}(X_i)|X_i)} - 1 \right] \{ q_{\tau_{0,K_0}(X_i)-1,0,K_0(X_i)}(X_i) - \hat{q}_{\tau_{0,K_0}(X_i)-1,0,K_0(X_i)}(X_i) \} \].

where

\[ \hat{V}_1 = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\mathbb{I}\{A_i \in \hat{d}(X_i)\}}{\hat{e}(\hat{d}(X_i)|X_i)} \{ Y_i - q_{\tau_{0,K_0}(X_i)-1,0,K_0(X_i)}(X_i) \} + q_{\tau_{0,K_0}(X_i)-1,0,K_0(X_i)}(X_i) \].
In Part 1, we first establish the following result that

$$\eta_7 = o_p(n^{-1/2}).$$

(A.112)

This implies

$$\hat{\nu} = \hat{\nu}_1 + o_p(n^{-1/2}).$$

(A.113)

In the second step, we further decompose $\hat{\nu}_1$ as

$$\hat{\nu}_1 = \hat{\nu}_2 + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\{A_i \in \hat{d}(X_i)\} \{Y_i - q_{s,0}(X_i)\}}{e(d_0(X_i)|X_i)} - \frac{\{A_i \in d_0(X_i)\}}{e(d_0(X_i)|X_i)} \right\} \left\{ Y_i - q_{s,0}(X_i) \right\},$$

where

$$\hat{\nu}_2 = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\{A_i \in d_0(X_i)\} \{Y_i - q_{s,0}(X_i)\}}{e(d_0(X_i)|X_i)} \right\} \left\{ Y_i - q_{s,0}(X_i) \right\}.$$

We focus on proving

$$\eta_8 = o_p(n^{-1/2}).$$

(A.114)

This together with (A.113) leads to

$$\hat{\nu} = \hat{\nu}_2 + o_p(n^{-1/2}).$$

(A.115)

Combing the results in the first two steps, it follows from the definition of $d_0(\cdot)$ that

$$\hat{\nu} = \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{S} \in \mathcal{P}_0} \left\{ \mathbb{I}(\mathcal{S} = d_0(X_i)) \frac{\{A_i \in d_0(X_i)\}}{e(d_0(X_i)|X_i)} \left\{ Y_i - q_{s,0}(X_i) \right\} + q_{s,0}(X_i) \right\} + o_p(n^{-1/2}),$$

almost surely. Notice that the first term at RHS corresponds to a sum of i.i.d random variables. Hence, based on Lindeberg-Feller central limit theorem, one can show the asymptotic normality result of the value estimator under the proposed I2DR.
Proof of Part 1: We aim to show (A.112). Toward that end, we define

\[
\hat{V}_3 = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\mathbb{I}[A_i \in \hat{d}(X_i)]}{e(\hat{d}(X_i)|X_i)} \left( Y_i - q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) \right) + q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) \right].
\]

The difference \(|\eta_3|\) can be upper bounded by \(|\hat{V}_1 - \hat{V}_3| + |\hat{V} - \hat{V}_3|\). Consider \(|\hat{V}_1 - \hat{V}_3|\) first. Under the given conditions, the term \(\left\{ \frac{\mathbb{I}[A_i \in \hat{d}(X_i)]}{e(\hat{d}(X_i)|X_i)} - 1 \right\} q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i)\) is bounded, it suffices to show that

\[
\frac{1}{n} \sum_{i=1}^{n} \left| q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) - q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) \right| = o_p(n^{-1/2}),
\]

where \(K_0()\) and \(\tilde{K}(\cdot)\) are defined in (A.104) and (A.99), respectively. Under the margin-type condition, the above expression can be proven using similar arguments in the proof of Theorem 4. We omit the details to save space.

It remains to show \(|\hat{V} - \hat{V}_3| = o_p(n^{1/2})\). Notice that \(|\hat{V} - \hat{V}_3|\) can be further upper bounded by

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}[A_i \in \hat{d}(X_i)]}{e(\hat{d}(X_i)|X_i)} \left\{ q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) - q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) \right\} + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\mathbb{I}[A_i \in \hat{d}(X_i)]}{e(\hat{d}(X_i)|X_i)} - \frac{\mathbb{I}[A_i \in \hat{d}(X_i)]}{e(\hat{d}(X_i)|X_i)} \right\} \left\{ q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) - q_{\tau_0,\hat{R}(X_i),1-\tau_0,\hat{R}(X_i)}(X_i) \right\}.
\]

Consider the first line. Notice that it can be represented by

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{G} \in \mathcal{F}} \left\{ \frac{\mathbb{I}[A_i \in \mathcal{G}]}{e(\mathcal{G}|X_i)} - 1 \right\} \left\{ q_{\mathcal{G},0}(X_i) - q_{\mathcal{G}}(X_i) \right\} \mathbb{I}(\mathcal{G} = \hat{d}(X_i)).
\]

Since the number of intervals in \(\mathcal{F}\) is finite with probability tending to 1 (see Results (i) in Theorem 1), to show the above expression is \(o_p(n^{-1/2})\), it suffices to show

\[
\sup_{\mathcal{F} \in \mathcal{F}(m)} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\mathbb{I}[A_i \in \mathcal{G}]}{e(\mathcal{G}|X_i)} - 1 \right\} \left\{ q_{\mathcal{G},0}(X_i) - q_{\mathcal{G}}(X_i) \right\} \mathbb{I}(\mathcal{G} = \hat{d}(X_i)) \right| = o_p(n^{-1/2}).
\]

The key observation is that, by Corollary A.1 of Chernozhukov et al. (2014), the above empirical sum forms a VC-type class. Using similar arguments in bounding the stochastic error in Step 2 of the proof of Lemma 5, we can show the above assertion holds.

To bound the second line, notice that by Cauchy-Schwarz inequality, it is smaller than
or equal to the square root of
\[
\frac{1}{n} \sum_{i=1}^{n} \left| \frac{\mathbb{I}\{A_i \in \tilde{d}(X_i)\}}{e(\tilde{d}(X_i)|X_i)} - \frac{\mathbb{I}\{A_i \in \tilde{d}(X_i)\}}{\tilde{e}(\tilde{d}(X_i)|X_i)} \right|^2 - \frac{1}{n} \sum_{i=1}^{n} \left| q_{\tilde{\tau}_{0,R(X_i)|-,\tau_0,R(X_i)},0}(X_i) - \tilde{q}_{\tilde{\tau}_{0,R(X_i)|-,\tau_0,R(X_i)}}(X_i) \right|^2
\]

Using similar arguments in establishing the uniform convergence rate of \( \tilde{q} \), we can show that \( \eta_7^{(2)} = o_p(n^{-c}) \) for some \( c > 1/2 \). To prove the second line is \( o_p(n^{-1/2}) \), it remains to show \( \eta_7^{(1)} = O_p(n^{-1/2} \log n) \). Under the positivity assumption on \( e \) and \( \tilde{e} \), it suffices to show

\[
\frac{1}{n} \sum_{i=1}^{n} \left| e(\tilde{d}(X_i)|X_i) - \tilde{e}(\tilde{d}(X_i)|X_i) \right|^2 = O_p(n^{-1/2} \log n).
\] (A.116)

The left-hand-side can be further upper bounded by

\[
\frac{1}{n} \sum_{\mathcal{S} \in \mathcal{D}} \sum_{i=1}^{n} \left| e(\mathcal{S}|X_i) - \tilde{e}(\mathcal{S}|X_i) \right|^2 \leq \sum_{\mathcal{S} \in \mathcal{D}} E|e(\mathcal{S}|X) - \tilde{e}(\mathcal{S}|X)|^2 + \sum_{\mathcal{S} \in \mathcal{D}} \left[ \frac{1}{n} \sum_{i=1}^{n} \left| e(\mathcal{S}|X_i) - \tilde{e}(\mathcal{S}|X_i) \right|^2 - E|e(\mathcal{S}|X) - \tilde{e}(\mathcal{S}|X)|^2 \right].
\]

The first term on the second line is \( O_p(n^{-1/2}) \) under Condition (A8) and the fact that \( |\mathcal{D}| = O(1) \) with probability tending to 1. To prove (A.116), by the boundedness of \( |\mathcal{D}| \), it suffices to show the supremum of the empirical process term

\[
\sup_{\mathcal{S} \in \mathcal{D}(m)} \left[ \frac{1}{n} \sum_{i=1}^{n} \left| e(\mathcal{S}|X_i) - \tilde{e}(\mathcal{S}|X_i) \right|^2 - E|e(\mathcal{S}|X) - \tilde{e}(\mathcal{S}|X)|^2 \right] = O_p(n^{-1/2} \log n).
\]

Under Condition (A8), this can be proven in a similar manner as Step 2 of the proof of Lemma 5. We omit the details to save space.

**Proof of Part 2:** We next focus on proving (A.114). We notice that \( |\eta_8| \) can be upper bounded by

\[
\left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\mathbb{I}\{A_i \in \tilde{d}(X_i)\}}{e(\tilde{d}(X_i)|X_i)} - \frac{\mathbb{I}\{A_i \in \tilde{d}(X_i)\}}{e(\tilde{d}(X_i)|X_i)} \right\} \left\{ Y_i - q_{\tilde{\tau}_{0,\mathbb{E}(X_i)|-,\tau_0,\mathbb{E}(X_i)},0}(X_i) \right\} \right| + \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\mathbb{I}\{A_i \in d_0(X_i)\}}{e(d_0(X_i)|X_i)} - \frac{\mathbb{I}\{A_i \in d_0(X_i)\}}{e(\tilde{d}(X_i)|X_i)} \right\} \left\{ Y_i - q_{\tilde{\tau}_{0,\mathbb{E}(X_i)|-,\tau_0,\mathbb{E}(X_i)},0}(X_i) \right\} \right|.
\]
The first line can be shown to be \( o_p(n^{-1/2}) \) using similar arguments in the proof of Part 1. The second line can be shown to be \( o_p(n^{1/2}) \) by noting that the difference between \( d_0 \) and \( \hat{d} \) is asymptotically negligible. This completes the proof.

### A.2.13 Proof of Theorem 6

Before proving Theorem 6, it is worth mentioning that results in Lemma 1 and Lemma 4 do not rely on the assumption that \( \theta_0(\cdot) \) is piecewise constant. These lemmas hold under the conditions in Theorem 6 as well. The proof is divided into two parts. In the first part, we derive the convergence rate of the integrated \( \ell_2 \) loss for \( b_{\theta} \). Then, we establish the convergence rate of the value under our I2DR.

**Convergence rate of the integrated \( \ell_2 \) loss:** We first establish the upper error bound on the integrated \( \ell_2 \) loss of \( b_{\theta} \). Here, we consider a more general framework. Specifically, define

\[
AE_k(\theta_0) = \inf_{(\theta_{\mathcal{P}}, \mathcal{I}) \in \mathcal{P} \colon |\mathcal{P}| \leq k+1} \left\{ \sup_{a \in [0,1]} \left\| \theta_0(a) - \sum_{\mathcal{I} \in \mathcal{P}} \theta_{\mathcal{I}} \mathbb{1}(a \in \mathcal{I}) \right\|_2 \right\}.
\]

It describes how well \( \theta_0(\cdot) \) can be approximated by a step function with at most \( k \) change points. Consider the following class of functions

\[
\mathbb{B}^{a_0} = \left\{ \theta_0(\cdot) : \limsup_{k \to \infty} k^{a_0} AE_k(\theta_0) < \infty \right\},
\]

for some \( a_0 > 0 \). The parameter \( a_0 \) characterizes the speed of approximation as the number of change points increases. According to the discussion in Section 2.4.2, the class of Hölder continuous functions in Model II belongs to \( \mathbb{B}^{a_0} \). In the following, we show with probability at least \( 1 - O(n^{-2}) \) that \( \int_0^1 \| \hat{\theta}(a) - \theta_0(a) \|_2^2 \, da \leq \tilde{c}_{\gamma} \gamma_{2a_0/(1+2a_0)}^2 n \) for any \( \theta_0(\cdot) \in \mathbb{B}^{a_0} \).

Since \( \theta_0(\cdot) \in \mathbb{B}^{a_0} \), for some sequence \( \{k_n\}_{n} \) that satisfies \( k_n \to \infty \) as \( n \to \infty \), there exists a piecewise constant function \( \theta^*(\cdot) \) such that

\[
\theta^*(a) = \sum_{\mathcal{I} \in \mathcal{P}^*} \theta_{\mathcal{I}}^* \mathbb{1}(a \in \mathcal{I}), \quad \forall a \in [0,1],
\]

for some partition \( \mathcal{P}^* \) of \([0,1]\) with \( |\mathcal{P}^*| \leq k_n + 1 \) and some \( (\theta_{\mathcal{I}}^*)_{\mathcal{I} \in \mathcal{P}^*} \in \prod_{\mathcal{I} \in \mathcal{P}^*} \mathbb{R}^{p+1} \), and

\[
\sup_{\mathcal{I} \in \mathcal{P}^*} \sup_{a \in \mathcal{I}} \| \theta_0(a) - \theta_{\mathcal{I}}^* \|_2 \leq \frac{c_4}{k_n^{a_0}}, \tag{A.117}
\]
for some constant \( c_4 > 0 \). Detailed choice of \( k_n \) will be given later. Combining (A.117) together with (A.12), we obtain that

\[
\sup_{I \in \mathcal{P}} \| \theta^*_I \|_2 \leq 2c_0, \tag{A.118}
\]

for sufficiently large \( n \).

Let \( \{ \tau^*_k \}_{k=1}^{|\mathcal{P}^*|-1} \) with \( 0 < \tau^*_1 < \tau^*_2 < \cdots < \tau^*_{|\mathcal{P}^*|-1} < 1 \) be the locations of the change points in \( J(\mathcal{P}^*) \). For \( 1 \leq k \leq |\mathcal{P}^*|-1 \), define \( \tau_k^* \) such that \( 0 \leq \tau_k^* - \tau_k^* N < 1/m \) and \( \tau_k^* \in \{ 1/m, 2/m, \ldots, 1 \} \). Let \( k_n^* \) be the largest integer that satisfies \( k_n^* \leq |\mathcal{P}^*| - 1 \) and \( \tau_k^* N < 1 \). Apparently, \( k_n^* \leq k_n \). Set \( \tau_0^* = \tau_0^* = 0 \) and \( \tau_{k_n^*+1}^* = \tau_{k_n^*+1}^* = 1 \). Define a new partition \( \mathcal{P}^{**} \in \mathcal{B}(m) \) and the set of vectors \( (\theta^{**}_I)_{I \in \mathcal{P}^{**}} \) as follows,

\[
\mathcal{P}^{**} = \{ [\tau_0^{**}, \tau_1^{**}), [\tau_1^{**}, \tau_2^{**}), \ldots, [\tau_{k_n^*}^{**}, \tau_{k_n^*+1}^{**}) \}, \quad \theta^{**}_{[\tau_k^{**}, \tau_{k+1}^{**})} = \theta^*_{[\tau_k^{**}, \tau_{k+1}^{**})}, \quad \forall k \in \{ 0, 1, \ldots, k_n^*-1 \} \text{ and } \theta^{**}_{[\tau_{k_n^*}^{**}, 1]} = \theta^*_{[\tau_{k_n^*}^{**}, 1]} \quad \text{(or } \theta^{**}_{[\tau_{k_n^*}^{**}, 1]} \text{).}
\]

Notice that it is possible that \( [\tau_k^{**}, \tau_{k+1}^{**}) = \emptyset \) for some \( k < k_n^* \).

Then, it follows from (A.118) that

\[
\sup_{I \in \mathcal{P}^{**}} \| \theta^*_I \|_2 \leq 2c_0, \tag{A.119}
\]

Moreover, it follows from (A.12), (A.119) and the condition \( m \asymp n \) that

\[
\sum_{I \in \mathcal{P}^{**}} \int_{\theta} \| \theta_0(a) - \theta^*_I \|_2^2 \, da \leq \sum_{I \in \mathcal{P}^{**}} \int_{\theta} \| \theta_0(a) - \theta^*_I \|_2^2 \, da + \frac{|\mathcal{P}^{**}|}{m} \sup_{a \in [0,1], I \in \mathcal{P}^{**}} \| \theta_0(a) - \theta^*_I \|_2^2 \leq c_4^2 k_n^{-2} + 9c_0^2 (k_n + 1) m^{-1} \leq O(1)(k_n^{-2} + n^{-1} k_n), \tag{A.120}
\]

for sufficiently large \( n \), where \( O(1) \) denotes some positive constant.
Notice that
\[
\sum_{i=1}^{n} \sum_{\mathcal{g} \in \mathcal{D}} \mathbb{I}(A_i \in \mathcal{g})(Y_i - \overline{X}_i \hat{\theta}_g)^2 = \sum_{i=1}^{n} \sum_{\mathcal{g}_1, \mathcal{g}_2 \in \mathcal{P}^*} \mathbb{I}(A_i \in \mathcal{g}_1 \cap \mathcal{g}_2)(Y_i - \overline{X}_i \hat{\theta}_{\mathcal{g}_1})^2
\]
\[
= \sum_{i=1}^{n} \sum_{\mathcal{g}_2 \in \mathcal{P}^*} \mathbb{I}(A_i \in \mathcal{g}_2)(Y_i - \overline{X}_i \hat{\theta}_{\mathcal{g}_2} + \overline{X}_i \bar{\theta} - \overline{X}_i \bar{\theta})^2
\]
\[
+ 2 \sum_{i=1}^{n} \sum_{\mathcal{g}_1, \mathcal{g}_2 \in \mathcal{P}^*} \mathbb{I}(A_i \in \mathcal{g}_1 \cap \mathcal{g}_2)(Y_i - \overline{X}_i \hat{\theta}_{\mathcal{g}_2})\overline{X}_i (\theta_{\mathcal{g}_2} - \overline{\theta}_{\mathcal{g}_2}).
\]

By definition, we have
\[
\sum_{i=1}^{n} \sum_{\mathcal{g} \in \mathcal{D}} \mathbb{I}(A_i \in \mathcal{g})(Y_i - \overline{X}_i \hat{\theta}_g)^2 + n\gamma_n|\mathcal{D}| \leq \sum_{i=1}^{n} \sum_{\mathcal{g} \in \mathcal{P}^*} \mathbb{I}(A_i \in \mathcal{g})(Y_i - \overline{X}_i \hat{\theta}_g)^2 + n(k_n + 1)\gamma_n.
\]
It follows that
\[
\chi_5 + 2\chi_6 + n\gamma_n |\mathcal{D}| \leq n(k_n + 1)\gamma_n. \tag{A.121}
\]

We now give a lower bound for \(\chi_5\). Similar to (A.40) and (A.41), we can show that the following event occurs with probability at least \(1 - O(n^{-2})\):
\[
\lambda_{\min} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{g})\overline{X}_i \overline{X}_i^\top \right) \geq c_5 n|\mathcal{g}|, \tag{A.122}
\]
for some constant \(c_5 > 0\), and any interval \(\mathcal{g} \in \mathcal{I}(m)\) that satisfies \(|\mathcal{g}| \geq \bar{c}_0 n^{-1} \log n\) where the constant \(\bar{c}_0\) is defined in Lemma 1. Under the event defined in (A.122), we obtain that
\[
\chi_5 \geq \sum_{\mathcal{g}_1, \mathcal{g}_2 \in \mathcal{P}^*} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{g}_1 \cap \mathcal{g}_2)\mathbb{I}(|\mathcal{g}_1 \cap \mathcal{g}_2| \geq \bar{c}_0 n^{-1} \log n)(\overline{X}_i \bar{\theta}_{\mathcal{g}_2} - \overline{X}_i \bar{\theta}_{\mathcal{g}_1})^2
\]
\[
\geq c_5 n \sum_{\mathcal{g}_1, \mathcal{g}_2 \in \mathcal{P}^*} \mathbb{I}(|\mathcal{g}_1 \cap \mathcal{g}_2| \geq \bar{c}_0 n^{-1} \log n)|\mathcal{g}_1 \cap \mathcal{g}_2| ||\bar{\theta}_{\mathcal{g}_2} - \bar{\theta}_{\mathcal{g}_1}||^2. \tag{A.123}
\]
In addition, under the events defined in (A.7) and Lemma 4, we have

\[ \sup_{\mathcal{I} \in \mathcal{P}} \| \hat{\theta}_x - \theta_{0,x} \|_2 \leq \sup_{\mathcal{I} \in \mathcal{P}} \frac{c_0 \sqrt{\log n}}{\sqrt{|\mathcal{I}| n}} \leq \frac{c_0 \sqrt{\log n}}{\sqrt{c_3 n \gamma_n}} = o(1), \]

since \( \gamma_n \gg n^{-1} \log n \). In view of (A.12), we obtain that \[ \sup_{\mathcal{I} \in \mathcal{P}} \| \hat{\theta}_x \|_2 \leq 2c_0, \quad (A.124) \]

for sufficiently large \( n \). This together with (A.119) yields that

\[
\sum_{\mathcal{I}_1 \in \mathcal{P}} \sum_{\mathcal{I}_2 \in \mathcal{P}^{*}} \mathbb{I}(|\mathcal{I}_1 \cap \mathcal{I}_2| \leq \bar{c}_0 n^{-1} \log n) |\mathcal{I}_1 \cap \mathcal{I}_2| \| \bar{\theta}_x - \hat{\theta}_{x_1} \|_2^2 \\
\leq (4c_0^2 \bar{c}_0 n^{-1} \log n) \sum_{\mathcal{I}_1 \in \mathcal{P}} \sum_{\mathcal{I}_2 \in \mathcal{P}^{*}} \mathbb{I}(|\mathcal{I}_1 \cap \mathcal{I}_2| \leq \bar{c}_0 n^{-1} \log n),
\]

with probability at least \( 1 - O(n^{-2}) \). Recall that \( \mathcal{P}^{*} \) has at most \( k_n \) change points. The number of nonempty intervals \( \mathcal{I}_1 \cap \mathcal{I}_2 \) is at most \( k_n + 1 + |\mathcal{P}| \). Thus, we obtain that

\[
\sum_{\mathcal{I}_1 \in \mathcal{P}} \sum_{\mathcal{I}_2 \in \mathcal{P}^{*}} \mathbb{I}(|\mathcal{I}_1 \cap \mathcal{I}_2| \leq \bar{c}_0 n^{-1} \log n) |\mathcal{I}_1 \cap \mathcal{I}_2| \| \theta_{\mathcal{P}}^{**} - \hat{\theta}_{x_1} \|_2^2 \\
\leq (k_n + 1 + |\mathcal{P}|)(4c_0^2 \bar{c}_0 n^{-1} \log n),
\]

with probability at least \( 1 - O(n^{-2}) \). This together with (A.123) yields that

\[
\chi_5 \geq c_5 n \sum_{\mathcal{I}_1 \in \mathcal{P}} \sum_{\mathcal{I}_2 \in \mathcal{P}^{*}} |\mathcal{I}_1 \cap \mathcal{I}_2| \| \bar{\theta}_x - \hat{\theta}_{x_1} \|_2^2 - c_5(k_n + 1 + |\mathcal{P}|)(4c_0^2 \bar{c}_0 \log n),
\]

with probability at least \( 1 - O(n^{-2}) \), or equivalently,

\[
\chi_5 \geq c_5 n \int_0^1 \| \hat{\theta}(a) - \theta^{**}(a) \|_2^2 da - c_5(k_n + 1 + |\mathcal{P}|)(4c_0^2 \bar{c}_0 \log n),
\quad (A.125)
\]

with probability at least \( 1 - O(n^{-2}) \), where

\[
\theta^{**}(a) = \sum_{\mathcal{I} \in \mathcal{P}^{*}} \theta^{**}_\mathcal{I} \mathbb{I}(a \in \mathcal{I}).
\]
We now provide an upper bound for $|\chi_6|$. Notice that

$$
\chi_6 = \sum_{i=1}^{n} \sum_{I_i \in \mathcal{I}_1} \sum_{J_i \in \mathcal{J}_2} \mathbb{I}(A_i \in \mathcal{I}_1 \cap \mathcal{J}_2) (X_i - \bar{X}_i^\top \theta_{\mathcal{J}_2}) (\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2})
$$

(A.126)

$$
= \sum_{i=1}^{n} \sum_{I_i \in \mathcal{I}_1} \sum_{J_i \in \mathcal{J}_2} \mathbb{I}(A_i \in \mathcal{I}_1 \cap \mathcal{J}_2) (Y_i - \bar{X}_i^\top \theta_0(A_i)) \bar{X}_i^\top (\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2})
$$

$$
+ \sum_{i=1}^{n} \sum_{I_i \in \mathcal{I}_1} \sum_{J_i \in \mathcal{J}_2} \mathbb{I}(A_i \in \mathcal{I}_1 \cap \mathcal{J}_2) (\bar{X}_i^\top \theta_0(A_i) - \bar{X}_i^\top \theta_{\mathcal{J}_2}^{**}) \bar{X}_i^\top (\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2}).
$$

It suffices to provide upper bounds for $|\chi_7|$ and $|\chi_8|$.

Under the event defined in (A.8), we obtain that

$$
\left| \sum_{i=1}^{n} \sum_{I_i \in \mathcal{I}_1} \sum_{J_i \in \mathcal{J}_2} \mathbb{I}(A_i \in \mathcal{I}_1 \cap \mathcal{J}_2) \mathbb{I}(|\mathcal{I}_1 \cap \mathcal{J}_2| \geq \tilde{c}_0 n^{-1} \log n) \{Y_i - \bar{X}_i^\top \theta_0(A_i)\} \bar{X}_i^\top (\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2}) \right|
$$

$$
\leq \sum_{I_1 \in \mathcal{I}_1} \sum_{J_2 \in \mathcal{J}_2} \left| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}_1 \cap \mathcal{J}_2) \mathbb{I}(|\mathcal{I}_1 \cap \mathcal{J}_2| \geq \tilde{c}_0 n^{-1} \log n) \{Y_i - \bar{X}_i^\top \theta_0(A_i)\} \bar{X}_i^\top \right| \|\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2}\|_2
$$

$$
\leq \sum_{I_1 \in \mathcal{I}_1} \sum_{J_2 \in \mathcal{J}_2} \sqrt{c_0 |\mathcal{I}_1 \cap \mathcal{J}_2| n \log n} \|\theta_{\mathcal{J}_2}^{**} - \hat{\theta}_{\mathcal{J}_2}\|_2 \leq \frac{c_5 n}{16} \int_0^1 \|\hat{\theta}(a) - \theta_{\mathcal{J}_2}^{**}(a)\|_2^2 da
$$

$$
+ \frac{4c_0 \log n}{c_5} \sum_{I_1 \in \mathcal{I}_1} \sum_{J_2 \in \mathcal{J}_2} |\mathcal{I}_1 \cap \mathcal{J}_2| \leq \frac{c_5 n}{16} \int_0^1 \|\hat{\theta}(a) - \theta_{\mathcal{J}_2}^{**}(a)\|_2^2 da + 4c_0 c_5^{-1} \log n,
$$

where the third inequality is due to Cauchy-Schwarz inequality.

In addition, using similar arguments in (A.22) and (A.23), we have with probability at least $1 - O(n^{-2})$ that, for any interval $\mathcal{I} \in \mathcal{I}(m)$ that satisfies $|\mathcal{I}| \leq \tilde{c}_0 n^{-1} \log n$,

$$
\left\| \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) \{Y_i - \bar{X}_i^\top \theta_0(A_i)\} \bar{X}_i \right\|_2 \leq \tilde{c}_3 \log n,
$$

(A.127)

for some constant $\tilde{c}_3 > 0$. Since the number of nonempty intervals $\mathcal{I}_1 \cap \mathcal{J}_2$ is at most
$k_n + 1 + |\mathcal{T}|$, we obtain that
\[
\left| \sum_{i=1}^{n} \sum_{J_i \in \mathcal{I}} \sum_{J_2 \in \mathcal{J}^*} \mathbb{I}(A_i \in J_1 \cap J_2) \| Y_i - \bar{X}_i^\top \theta_0(A_i) \|_2 \left( \theta_{J_2}^{**} - \hat{\theta}_J \right) \right| \\
\leq \sum_{J_1 \in \mathcal{I}} \sum_{J_2 \in \mathcal{J}^*} \left| \sum_{i=1}^{n} \mathbb{I}(A_i \in J_1 \cap J_2) \| Y_i - \bar{X}_i^\top \theta_0(A_i) \|_2 \left( \theta_{J_2}^{**} - \hat{\theta}_J \right) \right| \\
\leq (k_n + 1 + |\mathcal{T}|)(\bar{c}_2 \log n) \sup_{J_1 \in \mathcal{I}} \sup_{J_2 \in \mathcal{J}^*} \| \theta_{J_2}^{**} - \hat{\theta}_J \|_2 \leq 4c_0(k_n + 1 + |\mathcal{T}|)(\bar{c}_2 \log n),
\]
with probability at least $1 - O(n^{-2})$. It follows that
\[
|\chi_1| \leq \frac{c_5 n}{16} \int_0^1 \| \hat{\theta}(a) - \theta^{**}(a) \|_2^2 da + 4c_0c_5^{-1} \log n + 4c_0(k_n + 1 + |\mathcal{T}|)(\bar{c}_2 \log n), (A.128)
\]
with probability at least $1 - O(n^{-2})$.

As for $|\chi_8|$, it follows from Cauchy-Schwarz inequality that
\[
|\chi_8| \leq \frac{1}{4} \sum_{i=1}^{n} \sum_{J_i \in \mathcal{I}} \sum_{J_2 \in \mathcal{J}^*} \mathbb{I}(A_i \in J_1 \cap J_2) \| \bar{X}_i^\top \theta_{J_2}^{**} - \bar{X}_i^\top \hat{\theta}_J \|_2^2 \\
+ \sum_{i=1}^{n} \sum_{J_i \in \mathcal{I}} \sum_{J_2 \in \mathcal{J}^*} \mathbb{I}(A_i \in J_1 \cap J_2) (\bar{X}_i^\top \theta_0(A_i) - \bar{X}_i^\top \theta_{J_2}^{**})^2 = \frac{\chi_5}{4} + \chi_9. \tag{A.129}
\]
Notice that
\[
\chi_9 = \sum_{i=1}^{n} \sum_{J_i \in \mathcal{I}} \mathbb{I}(A_i \in J_i) (\bar{X}_i^\top \theta_0(A_i) - \bar{X}_i^\top \theta_{J_i}^{**})^2
\]
It follows from (A.33), (A.35), (A.120) and Cauchy-Schwarz inequality that
\[
\mathbb{E}(\chi_9) = n \sum_{\mathcal{I} \in \mathcal{J}^*} \mathbb{E}(A \in \mathcal{I}) (\bar{X}_i^\top \theta_0(A) - \bar{X}_i^\top \theta_{J_i}^{**})^2 \leq n \sum_{\mathcal{I} \in \mathcal{J}^*} \mathbb{E}(\| \bar{X} \|_2^2 I(A \in \mathcal{I}) | \theta_0(A) - \theta_{J_i}^{**} \|_2^2 \\
\leq n \sum_{\mathcal{I} \in \mathcal{J}^*} \mathbb{E}(\| \bar{X} \|_2^2 | A) \| \theta_0(A) - \theta_{J_i}^{**} \|_2^2 \leq \omega^2 n \sum_{\mathcal{I} \in \mathcal{J}^*} \mathbb{E}(A \in \mathcal{I}) | \theta_0(A) - \theta_{J_i}^{**} \|_2^2 \\
\leq C_0 \omega^2 n \sum_{\mathcal{I} \in \mathcal{J}^*} \int \| \theta_0(a) - \theta_{J_i}^{**} \|_2^2 da \leq O(1)(n \kappa_n^{-2\alpha_0} + \kappa_n),
\]
where $O(1)$ denotes some positive constant. Using similar arguments in (A.53), we have for
any integer \( q \geq 2 \) that
\[
\mathbb{E}
\left(
\sum_{A \in \mathcal{F}} \mathbb{I}(A \in \mathcal{F}) \left[ X^\top \theta_0(A) - X^\top \theta^* \right]^2 \right)^q
\leq \sum_{A \in \mathcal{F}} \mathbb{E} \mathbb{I}(A \in \mathcal{F}) \left[ X^\top \theta_0(A) - X^\top \theta^* \right]^{2q}
\leq q! c^q \sum_{A \in \mathcal{F}} \int_{\mathcal{F}} \| \theta_0(a) - \theta^* \|_2^2 \, da \leq q! C^q (n \kappa_n^{-2\alpha_0} + \kappa_n),
\]
for some constants \( c, C > 0 \). Using Bernstein’s inequality, we have for any \( t > 0 \) that
\[
\Pr(\chi_9 \geq \mathbb{E}\chi_9 + t) \leq \exp \left( -\frac{t^2}{2 \, t C + 2C^2 (n \kappa_n^{-2\alpha_0} + \kappa_n)} \right).
\]

We will require the sequence \( \{k_n\} \) to satisfy \( k_n \gg \log n \). Set \( t_0 = 4C \sqrt{(n \kappa_n^{-2\alpha_0} + \kappa_n) \log n} \), we have
\[
\frac{t_0^2}{t_0 C + 2C^2 (n \kappa_n^{-2\alpha_0} + \kappa_n)} = \frac{8\sqrt{n \kappa_n^{-2\alpha_0} + \kappa_n \log n}}{2 \log n + \sqrt{n \kappa_n^{-2\alpha_0} + \kappa_n}} \geq 2 \log n,
\]
for sufficiently large \( n \). Therefore, we obtain with probability at least \( 1 - O(n^{-2}) \) that
\[
\chi_9 \leq O(1)(n \kappa_n^{-2\alpha_0} + \kappa_n) + 4C \sqrt{(n \kappa_n^{-2\alpha_0} + \kappa_n) \log n} = O(n \kappa_n^{-2\alpha_0} + \kappa_n).
\]

This together with (A.126), (A.128) and (A.129) yields that
\[
|\chi_5| \leq \frac{c_5 n}{16} \int_0^1 \| \hat{\theta}(a) - \theta^*(a) \|_2^2 \, da + c_6 \{(k_n + |\mathcal{T}|) \log n + n \kappa_n^{-2\alpha_0} \} + \frac{\chi_5}{4}.
\]
with probability at least \( 1 - O(n^{-2}) \), for some constant \( c_6 > 0 \) and sufficiently large \( n \). In view of (A.121) and (A.125), we obtain with probability at least \( 1 - O(n^{-2}) \) that \( \chi_5 \leq n \gamma_n (k_n + 1 - \mathcal{T}) + 2|\chi_6| \) and hence
\[
\frac{3c_5 n}{8} \int_0^1 \| \hat{\theta}(a) - \theta^*(a) \|_2^2 \, da \leq 2c_6 \{(k_n + |\mathcal{T}|) \log n + n \kappa_n^{-2\alpha_0} \} + n \gamma_n (k_n + 1 - \mathcal{T}). \tag{A.130}
\]

Suppose \( |\mathcal{T}| \geq 2k_n + 1 \). Under the event defined in (A.130), it follows from the condition
\(n\gamma_n \gg \log n\) that
\[
n\gamma_n(k_n + 1 - |\mathcal{P}|) + 2c_6(k_n + |\mathcal{P}|)\log n \leq 3c_6|\mathcal{P}|\log n - 2^{-1}n\gamma_n|\mathcal{P}| \leq 0,
\]
for sufficiently large \(n\), and hence
\[
\frac{3c_5n}{8} \int_0^1 \|\hat{\theta}(a) - \theta^*(a)\|_2^2da \leq 2c_6n k_n^{-2\alpha_0}.
\tag{A.131}
\]

Otherwise, suppose \(|\mathcal{P}| \leq 2k_n\). It follows from (A.130) that
\[
\frac{3c_5n}{8} \int_0^1 \|\hat{\theta}(a) - \theta^*(a)\|_2^2da \leq 6c_6(k_n \log n + nk_n^{-2\alpha_0}) + n\gamma_n k_n,
\]
with probability at least \(1 - O(n^{-2})\). This together with (A.131) yields that
\[
\int_0^1 \|\hat{\theta}(a) - \theta^*(a)\|_2^2da \leq 16c_5^{-1} c_6^{-1}(k_n \log n + nk_n^{-2\alpha_0}) + 3\gamma_n k_n,
\tag{A.132}
\]
with probability at least \(1 - O(n^{-2})\).

By Cauchy-Schwarz inequality, we have
\[
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2da = \int_0^1 \|\hat{\theta}(a) - \theta^*(a) + \theta^*(a) - \theta_0(a)\|_2^2da \\
\leq 2\int_0^1 \|\hat{\theta}(a) - \theta^*(a)\|_2^2da + 2\int_0^1 \|\theta^*(a) - \theta_0(a)\|_2^2da.
\]
In view of (A.120) and (A.132), we obtain that
\[
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2da = O(n^{-1}k_n \log n + k_n^{-2\alpha_0} + \gamma_n k_n) = O(k_n^{-2\alpha_0} + \gamma_n k_n),
\tag{A.133}
\]
with probability at least \(1 - O(n^{-2})\), where the last equality is due to the condition that \(\gamma_n \gg n^{-1}\log n\). Set \(k_n = \lfloor \gamma_n^{- (1 + 2\alpha_0)} \rfloor\) (the largest integer that is smaller than \(\gamma_n^{- (1 + 2\alpha_0)}\)), we obtain that
\[
\int_0^1 \|\hat{\theta}(a) - \theta_0(a)\|_2^2da = O(\gamma_n^{2\alpha_0/(1 + 2\alpha_0)}),
\]
with probability at least $1 - O(n^{-2})$. The proof is hence completed.

**Convergence rate of the value function:** To derive the convergence rate of the value function under the proposed I2DR, we introduce the following lemma.

**Lemma 8** Assume conditions in Theorem 6 hold. Then for any interval $I \in \mathcal{I}(m)$ with $|I| \geq \bar{c}_0 n^{-1} \log n$ and any interval $I' \in \mathcal{P}$ with $I \subseteq I'$, we have with probability at least $1 - O(n^{-2})$ that

$$
\|\theta_{0,I} - \theta_{0,I'}\|_2 \leq 3 c_5^{-1} \gamma_n |I|^{-1},
$$

where the constant $c_5$ is defined in (A.122).

Recall by the definition of the value function that

$$
V^{opt} - V^\pi(b) = E \left( \sup_{a \in [0,1]} X^\top \theta_0(a) \right) - E \left( \int_{\mathcal{D}(X)} X^\top \theta_0(a) \pi^*(a; X, \mathcal{D}(X)) \, da \right). \tag{A.134}
$$

We begin by providing an upper bound for

$$
\chi_{11} = E \left( \sup_{a \in [0,1]} X^\top \theta_0(a) \right) - E \left( \sup_{I \in \mathcal{P}} X^\top \theta_{0,I} \right).
$$

It follows from (A.33) and Cauchy-Schwarz inequality that

$$
\chi_{11} = E \left( \sup_{I \in \mathcal{P}} \sup_{a \in I} X^\top \theta_0(a) \right) - E \left( \sup_{I \in \mathcal{P}} X^\top \theta_{0,I} \right) \leq E\|X\|_2 \sup_{I \in \mathcal{P}} \sup_{a \in I} \|\theta_0(a) - \theta_0,I\|_2 \leq (p + 1)^{1/2} \omega \sup_{I \in \mathcal{P}} \sup_{a \in I} \|\theta_0(a) - \theta_0,I\|_2.
$$

Consider a sequence $\{d_n\}_n$ that satisfies $d_n \geq 0, \forall n$, $d_n \to 0$ as $n \to \infty$ and $d_n \gg n^{-1} \log n$. By the definition of Hölder continuous functions, we have for any $I$ with $|I| \leq d_n$ that

$$
\sup_{a_1, a_2 \in I} \|\theta_0(a_1) - \theta_0(a_2)\|_2 \leq L \sup_{a_1, a_2 \in I} |a_1 - a_2|^{\alpha_0} \leq L d_n^{\alpha_0}.
$$
It follows that
\[
\sup_{a \in \mathcal{I}} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2 \leq \sup_{a \in \mathcal{I}} \left\| \theta_0(a) - \{ E \tilde{X} \tilde{X}^\top (A \in \mathcal{I}) \}^{-1} E \tilde{X} \tilde{X}^\top \theta_0(A) \| (A \in \mathcal{I}) \right\|_2 \\
\leq \sup_{a \in \mathcal{I}} \left\| \{ E \tilde{X} \tilde{X}^\top (A \in \mathcal{I}) \}^{-1} E \tilde{X} \tilde{X}^\top (A \in \mathcal{I}) \{ \theta_0(a) - \theta_0(A) \} \right\|_2 \\
\leq \sup_{a \in \mathcal{I}} \left\| E \tilde{X} \tilde{X}^\top (A \in \mathcal{I}) \right\|_2 \sup_{a, a^* \in \mathcal{I}} \| \theta_0(a) - \theta_0(a^*) \|_2 \leq L d_n^{a_0}, \tag{A.136}
\]
for any $\mathcal{I}$ that satisfies $|\mathcal{I}| \leq d_n$.

Consider an interval $\mathcal{I} \in \mathcal{I}(m)$ that satisfies $|\mathcal{I}| > d_n$. For any $a \in \mathcal{I}$, we can find an interval $\mathcal{I}' \subseteq \mathcal{I}$ with $d_n/2 \leq |\mathcal{I}'| \leq d_n$ and $\mathcal{I}' \in \mathcal{I}(m)$ that covers $a$. Similar to (A.136), we have
\[
\| \theta_{0,\mathcal{I}''} - \theta_0(a) \|_2 \leq L d_n^{a_0}. \tag{A.137}
\]
Since $d_n \gg n^{-1} \log n$, by Lemma 8, we have with probability at least $1 - O(n^{-2})$ that
\[
\| \theta_{0,\mathcal{I}} - \theta_0(a) \|_2 \leq 3 \sqrt{2 c_5^{-1} \gamma_n d_n^{-1}}.
\]
This together with (A.137) yields that
\[
\sup_{a \in \mathcal{I}} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2 \leq L d_n^{a_0} + 3 \sqrt{2 c_5^{-1} \gamma_n d_n^{-1}},
\]
for any $\mathcal{I} \in \mathcal{I}(m)$ that satisfies $|\mathcal{I}| > d_n$. Combining this together with (A.136), we obtain that
\[
\sup_{a \in \mathcal{I}} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2 \leq L d_n^{a_0} + 3 \sqrt{2 c_5^{-1} \gamma_n d_n^{-1}},
\]
for any $\mathcal{I} \in \mathcal{I}(m)$, with probability at least $1 - O(n^{-2})$. Set $d_n \approx \gamma_n^{(1+2a_0)^{-1}}$, we have with probability at least $1 - O(n^{-2})$ that
\[
\sup_{a \in \mathcal{I}} \| \theta_0(a) - \theta_{0,\mathcal{I}} \|_2 \leq O(1) \gamma_n^{a_0/(1+2a_0)},
\]
for any $\mathcal{I} \in \mathcal{I}(m)$, where $O(1)$ denotes some positive constant.

Therefore, we obtain with probability at least $1 - O(n^{-2})$ that
\[
\chi_{11} \leq O(1) \gamma_n^{a_0/(1+2a_0)}, \tag{A.138}
\]

where \(O(1)\) denotes some positive constant. Similarly, we can show with probability at least \(1 - O(n^{-2})\) that

\[
\chi_{12} = \mathbb{E}X^\top \theta_{0, \hat{d}(X)} - \mathbb{E}\left( \int \hat{d}(X) \theta_0(a) \pi^*(a; X, \hat{d}(X)) da \right) \leq O(1)\gamma_n^{a_0/(1+2a_0)},
\]

where \(O(1)\) denotes some positive constant. This together with (A.134) and (A.138) yields that,

\[
V^{opt} - V^{\pi^*}(\hat{d}) \leq \mathbb{E}\left( \sup_{\mathcal{G} \in \mathcal{F}} \bar{X}^\top \theta_{0, \mathcal{G}} \right) - \mathbb{E}\left( \sup_{\mathcal{G} \in \mathcal{F}} \bar{X}^\top \hat{\theta}_{\mathcal{G}} \right) + O(1)\gamma_n^{a_0/(1+2a_0)},
\]  (A.139)

with probability at least \(1 - O(n^{-2})\), where \(O(1)\) denotes some positive constant.

Using similar arguments in (A.135), we can show that

\[
\mathbb{E}\left( \sup_{\mathcal{G} \in \mathcal{F}} \bar{X}^\top \theta_{0, \mathcal{G}} \right) - \mathbb{E}\left( \sup_{\mathcal{G} \in \mathcal{F}} \bar{X}^\top \hat{\theta}_{\mathcal{G}} \right) \leq (p + 1)^{1/2} \omega \sup_{\mathcal{G} \in \mathcal{F}} \|\theta_{0, \mathcal{G}} - \hat{\theta}_{\mathcal{G}}\|_2,
\]

and

\[
\mathbb{E}X^\top \theta_{0, \hat{d}(X)} - \mathbb{E}X^\top \hat{\theta}_{\hat{d}(X)} \leq (p + 1)^{1/2} \omega \sup_{\mathcal{G} \in \mathcal{F}} \|\theta_{0, \mathcal{G}} - \hat{\theta}_{\mathcal{G}}\|_2.
\]

Since \(\sup_{\mathcal{G} \in \mathcal{F}} \bar{X}^\top \hat{\theta}_{\mathcal{G}} = \bar{X}^\top \theta_{0, \hat{d}(X)}\), we have

\[
V^{opt} - V^{\pi^*}(\hat{d}) \leq 2(p + 1)^{1/2} \omega \sup_{\mathcal{G} \in \mathcal{F}} \|\theta_{0, \mathcal{G}} - \hat{\theta}_{\mathcal{G}}\|_2 + O(1)\gamma_n^{a_0/(1+2a_0)},
\]  (A.140)

under the event defined in (A.139). It follows from Lemma 1 and 4 that

\[
\sup_{\mathcal{G} \in \mathcal{F}} \|\theta_{0, \mathcal{G}} - \hat{\theta}_{\mathcal{G}}\|_2 \leq \frac{\sqrt{\log n}}{\sqrt{n}\gamma_n},
\]

with probability at least \(1 - O(n^{-2})\). This together with (A.140) yields that

\[
V^{opt} - V^{\pi^*}(\hat{d}) \leq O(1)\left(\gamma_n^{a_0/(1+2a_0)} + \frac{\sqrt{\log n}}{\sqrt{n}\gamma_n}\right),
\]

with probability at least \(1 - O(n^{-2})\), where \(O(1)\) denotes some positive constant. Set \(\gamma_n \asymp (n^{-1/2} \log^{1/2} n)^{(2a_0+1)/(4a_0+1)}\), we obtain that \(V^{opt} - V^{\pi^*}(\hat{d}) = O(n^{-a_0/(1+4a_0)} \log^{a_0/(1+4a_0)} n)\), with probability at least \(1 - O(n^{-2})\). The proof is hence completed.
A.2.14 Proof of Lemma 8

For a given interval $\mathcal{I}' \in \hat{\mathcal{D}}$, the set of intervals $\mathcal{I}$ considered in Lemma 8 can be classified into the following three categories.

Category 1: $\mathcal{I} = \mathcal{I}'$. Then it is immediate to see that $\|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{I}'}\|_2 = 0$ and the assertion automatically holds.

Category 2: There exists another interval $\mathcal{I}^* \in \mathcal{I}(m)$ that satisfies $\mathcal{I}' = \mathcal{I}^* \cup \mathcal{I}$. Notice that the partition $\hat{\mathcal{P}}^* = \hat{\mathcal{P}} \cup \{\mathcal{I}^*\} \cup \mathcal{I} - \{\mathcal{I}'\}$ also belongs to $\mathcal{B}(m)$. By definition, we have

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{I}_0 \in \hat{\mathcal{P}}} \Pi(A_i \in \mathcal{I}_0)(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}_0})^2 + \lambda_n |\mathcal{I}_0| ||\hat{\theta}_{\mathcal{I}_0}||_2^2 + \gamma_n \|\hat{\mathcal{P}}^*\| \geq \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{I}_0 \in \hat{\mathcal{P}}} \Pi(A_i \in \mathcal{I}_0)(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}_0})^2 + \lambda_n |\mathcal{I}_0| ||\hat{\theta}_{\mathcal{I}_0}||_2^2 + \gamma_n \|\hat{\mathcal{P}}\|,$$

and hence

$$\frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I})(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I}^*)(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}^*})^2 + \lambda_n |\mathcal{I}^*| \|\hat{\theta}_{\mathcal{I}^*}\|_2^2 - \gamma_n.$$

It follows from the definition of $\hat{\theta}_{\mathcal{I}}$ that

$$\frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I}^*)(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}^*})^2 + \lambda_n |\mathcal{I}^*| \|\hat{\theta}_{\mathcal{I}^*}\|_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I}^*)(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}^*})^2 + \lambda_n |\mathcal{I}^*| \|\hat{\theta}_{\mathcal{I}^*}\|_2^2.$$

Therefore, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I})(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 \quad \text{(A.141)}$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I})(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 - \gamma_n.$$

Category 3: There exist two intervals $\mathcal{I}^*, \mathcal{I}^{**} \in \mathcal{I}(m)$ that satisfy $\mathcal{I}' = \mathcal{I}^* \cup \mathcal{I} \cup \mathcal{I}^{**}$. Using similar arguments in proving (A.141), we can show that

$$\frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I})(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 \geq \frac{1}{n} \sum_{i=1}^{n} \Pi(A_i \in \mathcal{I})(Y_i - \bar{X}_i \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 - 2\gamma_n.$$
Hence, regardless of whether \( \mathcal{I} \) belongs to Category 2, or it belongs to Category 3, we have

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \hat{\theta}_\mathcal{I})^2 + \lambda_n |\mathcal{I}||\hat{\theta}_\mathcal{I}|_2^2 \\
\geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \hat{\theta}_\mathcal{I})^2 + \lambda_n |\mathcal{I}|\|\hat{\theta}_\mathcal{I}\|_2^2 - 2\gamma_n \\
\geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \hat{\theta}_\mathcal{I})^2 - 2\gamma_n. \tag{A.142}
\]

Notice that \(|\mathcal{I}'| \geq |\mathcal{I}|\). Under the event defined in (A.7), we obtain that

\[
\|\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'}\|_2 \leq c_0 \sqrt{\log n} \frac{\log n}{|\mathcal{I}| n}.
\]

Similar to (A.18), we can show the following event occurs with probability at least \(1 - O(n^{-2})\),

\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \theta_{0, \mathcal{I}})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'}) \right| \leq O(1)n^{-1} \log n, \tag{A.143}
\]

where \(O(1)\) denotes some positive constant. Similarly, using Cauchy-Schwarz inequality, we can show with probability at least \(1 - O(n^{-2})\) that

\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(\overline{X}_i \theta_{0, \mathcal{I}'})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'}) \right| \leq \frac{1}{4n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})((\overline{X}_i \theta_{0, \mathcal{I}'})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'})^2 + \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})((\overline{X}_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'})^2 \\
\leq \frac{1}{4n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})((\overline{X}_i \theta_{0, \mathcal{I}'})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'})^2 + O(1)n^{-1} \log n,
\]

where \(O(1)\) denotes some positive constant. This together with (A.143) yields

\[
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \overline{X}_i \theta_{0, \mathcal{I}})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'}) \right| \leq \frac{1}{4n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})((\overline{X}_i \theta_{0, \mathcal{I}'})X_i (\hat{\theta}_{\mathcal{I}'} - \theta_{0, \mathcal{I}'})^2 + O(1)n^{-1} \log n,
\]

with probability at least \(1 - O(n^{-2})\), where \(O(1)\) denote some positive constant. Using similar
arguments in proving (A.17), we can show the following event occurs with probability at least \(1 - O(n^{-2})\),

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \hat{\theta}_{\mathcal{I}})^2 \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 \tag{A.144}
\]

\[
- \frac{1}{2n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(\mathbf{X}_i^\top \theta_{0,\mathcal{I}} - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 - O(1)n^{-1} \log n,
\]

where \(O(1)\) denotes some positive constant.

In addition, it follows from the definition of \(\hat{\theta}_{\mathcal{I}}\) that

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\theta_{0,\mathcal{I}}\|_2^2.
\]

By (A.12) and the condition that \(\lambda_n = O(n^{-1} \log n)\), we obtain that

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \hat{\theta}_{\mathcal{I}})^2 + \lambda_n |\mathcal{I}| \|\hat{\theta}_{\mathcal{I}}\|_2^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 + O(1)n^{-1} \log n,
\]

where \(O(1)\) denotes some positive constant. This together with (A.142) and (A.144) yields

\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 \geq \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 - 2n\gamma_n - O(1) \log n
\]

\[
- \frac{1}{2n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(\mathbf{X}_i^\top \theta_{0,\mathcal{I}} - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2,
\]

with probability at least \(1 - O(n^{-2})\), where \(O(1)\) denotes some positive constant.

Notice that

\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 = \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}} + \mathbf{X}_i^\top \theta_{0,\mathcal{I}} - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2
\]

\[
= \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2 + 2 \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})(\mathbf{X}_i^\top \theta_{0,\mathcal{I}} - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})
\]

\[
+ \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I})(\mathbf{X}_i^\top \theta_{0,\mathcal{I}} - \mathbf{X}_i^\top \theta_{0,\mathcal{I}})^2.
\]
Combining this with (A.145) yields that
\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (\bar{X}_i \theta_{0,\mathcal{I}} - \bar{X}_i \theta_{0,\mathcal{J}})^2 \leq 4n\gamma_n + O(1) \log n + 4|\chi_{10}|. \tag{A.146}
\]

Under the event defined in (A.122), we obtain that
\[
\sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (\bar{X}_i \theta_{0,\mathcal{I}} - \bar{X}_i \theta_{0,\mathcal{J}})^2 \geq c_5 n |\mathcal{I}| \|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{J}}\|_2^2. \tag{A.147}
\]

By the definition of \(\theta_{0,\mathcal{I}}\), we have \(E_{\mathcal{I}}(A_i \in \mathcal{I}) (Y_i - \bar{X}_i \theta_{0,\mathcal{I}})X_i = 0\). Under the event defined in (A.8), it follows from Cauchy-Schwarz inequality that
\[
|\chi_{10}| \leq \frac{2}{c_5 n |\mathcal{I}|} \left( \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \bar{X}_i)X_i \right)^2 + \frac{c_5 n}{8} |\mathcal{I}| \|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{J}}\|_2^2 \leq \frac{2c_5^2 \log n}{c_5} + \frac{c_5 n}{8} |\mathcal{I}| \|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{J}}\|_2^2.
\]

This together with (A.146) and (A.147) yields that
\[
|\mathcal{I}| \|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{J}}\|_2^2 \leq \frac{8\gamma_n}{c_5} + O(1)n^{-1} \log n,
\]

with probability at least \(1 - O(n^{-2})\), where \(O(1)\) denotes some positive constant. Since \(\gamma_n \gg n^{-1} \log n\), for sufficiently large \(n\), we obtain with probability at least \(1 - O(n^{-2})\) that
\[
|\mathcal{I}| \|\theta_{0,\mathcal{I}} - \theta_{0,\mathcal{J}}\|_2^2 \leq 9c_5^{-1} \gamma_n.
\]

The proof is hence completed.

### A.2.15 Proof of Theorem 7

The proof of Theorem 7 is divided into three parts. In Part 1, we prove that for any interval \(\mathcal{I} \in \mathcal{I}(m)\) with \(|\mathcal{I}| \gg \gamma_n\) and any interval \(\mathcal{I}' \in \mathcal{P}c\) with \(\mathcal{I} \subseteq \mathcal{I}'\), we have with probability approaching 1 (w.p.a.1.) that
\[
E[q_{\mathcal{I},0}(X) - q_{\mathcal{I}',0}(X)]^2 \leq \tilde{C}|\mathcal{I}|^{-1} \gamma_n, \tag{A.148}
\]

for some constant \(\tilde{C} > 0\).
In the second part, we show assertion (i) in Theorem 7 holds. Finally, we present the proof for assertion (ii) in Theorem 7. It is worth mentioning that results in Lemma 5 and Lemma 7 do not rely on the assumption that $Q(\cdot)$ is piecewise function. These lemmas hold under the conditions in Theorem 7 as well.

*Proof of Part 1:* For a given interval $\mathcal{J}' \in \widehat{\mathcal{P}}$, the set of intervals $\mathcal{J}$ considered in (A.148) can be classified into the following three categories.

*Category 1:* $\mathcal{J} = \mathcal{J}'$. It is immediate to see that $q_{\mathcal{J},0} = q_{\mathcal{J}',0}$ and the assertion automatically holds.

*Category 2:* There exists another interval $\mathcal{J}^* \in \mathcal{I}(m)$ that satisfies $\mathcal{J}' = \mathcal{J}^* \cup \mathcal{J}$. Notice that the partition $\widehat{\mathcal{P}}^* = \widehat{\mathcal{P}} \cup \{\mathcal{J}^*\} \cup \mathcal{J} - \{\mathcal{J}'\}$ forms another partition. By definition, we have

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{J}_{0} \in \widehat{\mathcal{I}}_{0}} \mathbb{I}(A \in \mathcal{J}_{0}) \{Y_{i} - \tilde{q}_{\mathcal{J}_{0}}(X_{i})\}^2 + \gamma_{n} |\mathcal{P}| \geq \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{J}_{0} \in \widehat{\mathcal{I}}_{0}} \mathbb{I}(A \in \mathcal{J}_{0}) \{Y_{i} - \tilde{q}_{\mathcal{J}_{0}}(X_{i})\}^2 + \gamma_{n} |\mathcal{P}|,$$

and hence

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}) \{Y_{i} - \tilde{q}_{\mathcal{J}}(X_{i})\}^2 + \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}^*) \{Y_{i} - \tilde{q}_{\mathcal{J}^*}(X_{i})\}^2 \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}') \{Y_{i} - \tilde{q}_{\mathcal{J}'}(X_{i})\}^2 - \gamma_{n}.$$

It follows from the definition of $\tilde{q}_{\mathcal{J}^*}$ that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}^*) \{Y_{i} - \tilde{q}_{\mathcal{J}^*}(X_{i})\}^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}^*) \{Y_{i} - \tilde{q}_{\mathcal{J}^*}(X_{i})\}^2.$$

Therefore, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}) \{Y_{i} - \tilde{q}_{\mathcal{J}}(X_{i})\}^2 \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}) \{Y_{i} - \tilde{q}_{\mathcal{J}}(X_{i})\}^2 - \gamma_{n}. \quad (A.149)$$

*Category 3:* There exist two intervals $\mathcal{J}^*, \mathcal{J}^{**} \in \mathcal{I}(m)$ that satisfy $\mathcal{J}' = \mathcal{J}^* \cup \mathcal{J} \cup \mathcal{J}^{**}$. Using similar arguments in proving (A.149), we can show that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}) \{Y_{i} - \tilde{q}_{\mathcal{J}}(X_{i})\}^2 \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A \in \mathcal{J}) \{Y_{i} - \tilde{q}_{\mathcal{J}}(X_{i})\}^2 - 2\gamma_{n}.$$
Hence, regardless of whether \( \mathcal{I} \) belongs to Category 2, or it belongs to Category 3, we have

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \tilde{q}_\mathcal{I}_0(X_i))^2 \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \tilde{q}_\mathcal{I}(X_i))^2 - 2\gamma_n. \tag{A.150}
\]

Notice that for any interval \( \mathcal{I}_0 \),

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}_0) (Y_i - \tilde{q}_\mathcal{I}_0(X_i))^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}_0) (\tilde{q}_\mathcal{I}_0(X_i) - \tilde{q}_\mathcal{I}_0(X_i)) \{q_{\mathcal{I}_0,0}(X_i) - \tilde{q}_\mathcal{I}_0(X_i)\}^2
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}_0) (Y_i - \tilde{q}_\mathcal{I}_0(X_i))^2 - \mathbb{E}(A \in \mathcal{I}_0) (\tilde{q}_\mathcal{I}_0(X_i) - \tilde{q}_\mathcal{I}_0(X))^2.
\]

Using similar arguments in bounding the stochastic error term in Part 2 of the proof of Lemma 5, we can show w.p.a.1. that the RHS is of the order \( O(n^{-2\beta/(2\beta+p)} \log^8 n) \), for any \( \mathcal{I}_0 \in \mathcal{I}(m) \). As such, we obtain w.p.a.1. that

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \tilde{q}_\mathcal{I}(X_i))^2 = \mathbb{E}(A \in \mathcal{I}) (Y - \tilde{q}_\mathcal{I}(X))^2
\]

\[
+ O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta+p)} \log^8 n,
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \tilde{q}_\mathcal{I}_0(X_i))^2 = \mathbb{E}(A \in \mathcal{I}) (Y - \tilde{q}_\mathcal{I}(X))^2
\]

\[
+ O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta+p)} \log^8 n,
\]

where \( O(1) \) denotes some universal positive constant. Combining these together with (A.150) yields

\[
\mathbb{E}(A \in \mathcal{I}) (Y - \tilde{q}_\mathcal{I}(X))^2 \geq \mathbb{E}(A \in \mathcal{I}) (Y - \tilde{q}_\mathcal{I}_0(X))^2
\]

\[
- 2\gamma_n + O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta+p)} \log^8 n,
\]

for any \( \mathcal{I} \) and \( \mathcal{I}' \), w.p.a.1. Note that \( q_{\mathcal{I},0} \) satisfies \( \mathbb{E}(A \in \mathcal{I}) (Y - q_{\mathcal{I},0}(X)) \mid X = 0 \). We have

\[
\mathbb{E}(A \in \mathcal{I}) (q_{\mathcal{I},0}(X) - \tilde{q}_\mathcal{I}(X))^2 \geq \mathbb{E}(A \in \mathcal{I}) (q_{\mathcal{I},0}(X) - \tilde{q}_\mathcal{I}_0(X))^2
\]

\[
- 2\gamma_n + O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta+p)} \log^8 n.
\]
Consider the first term on the RHS. Note that

\[
E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - \tilde{q}_{x'}(X))^2] = E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - q_{x',0}(X))^2] + E[\mathbb{I}(A \in \mathcal{I}) (\tilde{q}_{x'}(X) - q_{x',0}(X))^2]
- 2E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - q_{x',0}(X)) (\tilde{q}_{x'}(X) - q_{x',0}(X))].
\]

By Cauchy-Schwarz inequality, the last term on the RHS can be lower bounded by

\[
-\frac{1}{2} E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - q_{x',0}(X))^2] - 2E[\mathbb{I}(A \in \mathcal{I}) (\tilde{q}_{x'}(X) - q_{x',0}(X))^2].
\]

It follows that

\[
E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - \tilde{q}_{x'}(X))^2] \geq \frac{1}{2} E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - q_{x',0}(X))^2]
- 3E[\mathbb{I}(A \in \mathcal{I}) (\tilde{q}_{x'}(X) - q_{x',0}(X))^2],
\]

and hence

\[
\frac{1}{2} E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - q_{x',0}(X))^2] - 2\gamma_n + O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta + p)} \log^8 n
\leq E[\mathbb{I}(A \in \mathcal{I}) (q_{x,0}(X) - \tilde{q}_{x'}(X))^2] + 3E[\mathbb{I}(A \in \mathcal{I}) (q_{x',0}(X) - \tilde{q}_{x'}(X))^2].
\]

By Lemma 5, Lemma 7 and the positivity assumption, the RHS is upper bounded by

\[O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(p+2\beta)} \log^8 n\] for some universal positive constant \(O(1)\), w.p.a.1. We obtain w.p.a.1. that

\[
E[\mathbb{I}(A \in \mathcal{I}) (q_{x}(X) - q_{x'}(X))^2] = 4\gamma_n + O(1) |\mathcal{I}| (n |\mathcal{I}|)^{-2\beta/(2\beta + p)} \log^8 n,
\]

uniformly for any \(\mathcal{I}\) and \(\mathcal{I}'\), or equivalently,

\[
E\left[\frac{e(\mathcal{I} \mid X)}{|\mathcal{I}|} (q_{x}(X) - q_{x'}(X))^2\right] = 4\gamma_n |\mathcal{I}| + O(1) (n |\mathcal{I}|)^{-2\beta/(2\beta + p)} \log^8 n.
\]

By the positivity assumption, we have w.p.a.1. that

\[
E[(q_{x}(X) - q_{x'}(X))^2] = O(\gamma_n |\mathcal{I}|^{-1}) + O((n |\mathcal{I}|)^{-2\beta/(2\beta + p)} \log^8 n),
\]

uniformly for any \(\mathcal{I}\) and \(\mathcal{I}'\). The proof of (A.148) is hence completed by noting that \(|\gamma_n| \ll |\mathcal{I}|\).
Proof of Part 2: Consider a sequence \( \{d_n\}_n \) such that \( d_n \to 0 \) and \( d_n \gg \gamma_n \). We aim to show
\[
\inf_{a \in I'} \inf_{\mathcal{F}' \in \mathcal{F}} E[|Q(X, a) - \hat{q}_{\mathcal{F}'}(X)|^2] = o_p(1).
\]
By Lemma 5, it suffices to show
\[
\inf_{a \in I'} \inf_{\mathcal{F}' \in \mathcal{F}} E[|Q(X, a) - q_{\mathcal{F}',0}(X)|^2] = o(1).
\]
Suppose \( |\mathcal{F}'| \geq d_n \). Then according to (A.148), we can find some \( \mathcal{F} \) such that \( |\mathcal{F}| \gg \gamma_n \) and \( a \in \mathcal{F} \subseteq \mathcal{F}' \),
\[
E[q_{\mathcal{F},0}(X) - q_{\mathcal{F}',0}(X)]^2 \leq \hat{C}|\mathcal{F}|^{-1} \gamma_n.
\]
Since \( |\mathcal{F}| \to 0 \) and \( a \in \mathcal{F} \), it follows from the uniform continuity of the outcome regression function that \( |q_{\mathcal{F},0}(X) - Q(X, a)| \to 0 \). The assertion thus follows.

Next, suppose \( |\mathcal{F}'| < d_n \). Then \( |\mathcal{F}'| \to 0 \) as well. It follows from the uniform continuity of the outcome regression function that
\[
\inf_{a \in \mathcal{F}, \mathcal{F}' \in \mathcal{F}} |q_{\mathcal{F}',0}(X) - Q(X, a)| \to 0.
\]
The assertion thus follows. This completes the proof for the result (i).

Proof of Part 3: This part follows the second part of the proof of Theorem 6. Recall by the definition of the value function that
\[
V^{o_p} - V^{\pi^*}(\tilde{d}) = E \left( \sup_{a \in [0,1]} Q(X, a) \right) - E \left( \int_{\tilde{d}(X)} Q(X, a) \pi^*(a; X, \tilde{d}(X)) \, da \right). \tag{A.151}
\]
Using similar arguments in (A.135), under the result (i), we can show that
\[
\chi_{13} = E \left( \sup_{a \in [0,1]} Q(X, a) \right) - E \left( \sup_{\mathcal{F}' \in \mathcal{F}} q_{\mathcal{F}',0}(X) \right) = o_p(1). \tag{A.152}
\]
Similarly, we have
\[
\chi_{14} = E Q(X, \tilde{d}(X)) - E \left( \int_{\tilde{d}(X)} Q(X, a) \pi^*(a; X, \tilde{d}(X)) \, da \right) = o_p(1).
\]

This together with (A.151) and (A.152) yields that,

\[ V^{opt} - V^{\pi^*}(d) \leq E\left( \sup_{\mathcal{F}} q_{\mathcal{S},0}(X) \right) - EQ(X, \hat{d}(X)) + o_p(1). \]  \hspace{1cm} (A.153)

Using similar arguments in (A.135), we can obtain that

\[ E\left( \sup_{\mathcal{F}} q_{\mathcal{S},0}(X) \right) - E\left( \sup_{\mathcal{F}} \hat{q}_s(X) \right) \leq \tilde{C}_5 \sup_{\mathcal{F}} E[|q_{\mathcal{S},0}(X) - \hat{q}_s(X)|^2], \]

for some constant \( \tilde{C}_5 \), and

\[ EQ(X, \hat{d}(X)) - EQ(X, d(X)) \leq \tilde{C}_6 \sup_{\mathcal{F}} E[|q_{\mathcal{S},0}(X) - \hat{q}_s(X)|^2], \]

for some constant \( \tilde{C}_6 \).

Since \( \sup_{\mathcal{F}} \hat{q}_s(X) = \hat{Q}(X, \hat{d}(X)) \), by Lemma 5 together with (A.153), we have

\[ V^{opt} - V^{\pi^*}(d) = o_p(1). \]  \hspace{1cm} (A.154)

The proof is hence completed.
B.1 Analysis of Computational Complexity of DJL

We analyze the computational complexity for the proposed method as follows. There are three main dominating parts of the computation: the adaptive discretization, the estimations of conditional mean function and the propensity score function, and the construction of the final value estimator.

First, for the adaptive discretization on the treatment space (the main part of DJL, see Algorithm 2 Part III.3), we use the pruned exact linear time (PELT) method in Killick et al. (2012b) to solve the dynamic programing. This step requires at least $O(m)$ computing steps and at most $O(m^2)$ steps (Friedrich et al. 2008). According to Theorem 3.2 in Killick et al. (2012b), the expected computational cost is $O(m)$.

Second, for each step in the linear complexity of adaptive discretization, we need to train the deep neural network for the conditional mean function and the propensity score function to calculate the cost function. Here, the time and space complexity of training a deep learning model varies depending on the actual architecture used. In our implementation, we employ the commonly used multilayer perceptron (MLP) to estimate the function
Q and the propensity score in each segment. Suppose we use the standard fully connected MLPs of width $w$ and depth $D$ with feedforward pass and back-propagation under total $e$ epochs. Then according to the complexity analysis of neural networks, the computational complexity of modeling the function $Q$ and the propensity score is $O\{2 \times ne(D - 1)w^2\}$.

For the last part, the construction of the final value estimator based on $\mathcal{L}$-fold cross fitting, which repeats the above two steps $\mathcal{L}$ times. Therefore, by putting the above results together, the total expected computational complexity of the proposed DJL is $O\{\mathcal{L} \times m \times 2 \times ne(D - 1)w^2\}$. Note that the computation for the last part (i.e., cross-fitting) can be easily implemented in parallel computing, and thus the total expected computational complexity of the proposed DJL can be reduced to $O\{m \times 2 \times ne(D - 1)w^2\}$. 

187
B.2  Additional Experimental Results

We include additional experimental results in this section. First, the number of initial intervals $m$ represents a trade-off between the estimation bias and the computational cost, as illustrated in Figure B.1. In practice, we recommend to set $m = n/10$. When $n$ is small, the performance of the resulting value estimator is not overly sensitive to the choice of $c$ as long as $c$ is not too large. See the left panel of Figure B.1 for details. When $n$ is large, we further investigate the computational capacity of the proposed method by setting $m = n/10$ for large sample sizes and report the corresponding computational time in Table B.2. We use Scenario 1 and consider the sample size chosen from $n \in \{1000, 2000, 5000, 10000\}$ for illustration. It turns out that such a choice of $c$ can still handle datasets with a few thousand observations. Here, we use parallel computing to process each fold, as our algorithm employs data splitting and cross-fitting. This largely facilitates the computation, leading to shorter computation time compared to those listed in Table B.1. Finally, when $n$ is extremely large, setting $m = n/10$ might be computationally intensive. In addition to parallel computing, there are some other techniques we can use to handle datasets with large sample size. For instance, in the change-point literature, Lu et al. (2017) proposed an intelligence sampling method to identify multiple change points with long time series data. Their method would not lose much statistical efficiency, but is much more computationally efficient. It is possible to adopt such an intelligence sampling method to our setting for adaptive discretization. This would enable our method to handle large datasets.

Table B.1:  The averaged computational cost (in minutes) under the proposed deep jump learning and three kernel-based methods for Scenario 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>$&lt; 1$</td>
<td>$&lt; 1$</td>
<td>365</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>3</td>
<td>$&lt; 1$</td>
<td>773</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>7</td>
<td>1</td>
<td>$&gt; 1440$ (24 hours)</td>
<td>$&lt; 1$</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>14</td>
<td>2</td>
<td>$&gt; 2880$ (48 hours)</td>
<td>$&lt; 1$</td>
</tr>
</tbody>
</table>
Table B.2: The averaged computational cost under the proposed deep jump learning for Scenario 1 with large sample settings.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>n = 1000</th>
<th>n = 2000</th>
<th>n = 5000</th>
<th>n = 10000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Computational time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.92 minutes</td>
<td>30.40 minutes</td>
<td>1.32 hours</td>
<td>2.86 hours</td>
</tr>
</tbody>
</table>

Table B.3: The absolute error and the standard deviation (in parentheses) of the estimated values under the optimal policy via the proposed deep jump learning and three kernel-based methods for Scenario 1 to 4.

<table>
<thead>
<tr>
<th>n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>Deep Jump Learning</td>
<td>0.445(0.381)</td>
<td>0.398(0.391)</td>
<td>0.253(0.269)</td>
</tr>
<tr>
<td>V = 1.33</td>
<td>SLOPE</td>
<td>0.392(0.377)</td>
<td>0.385(0.549)</td>
<td>0.329(0.400)</td>
</tr>
<tr>
<td></td>
<td>Kallus and Zhou (2018)</td>
<td>0.656(0.787)</td>
<td>0.848(0.799)</td>
<td>1.163(0.884)</td>
</tr>
<tr>
<td></td>
<td>Colangelo and Lee (2020)</td>
<td>1.285(1.230)</td>
<td>1.473(1.304)</td>
<td>1.826(1.463)</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Deep Jump Learning</td>
<td>0.696(0.376)</td>
<td>0.502(0.311)</td>
<td>0.400(0.219)</td>
</tr>
<tr>
<td>V = 1.00</td>
<td>SLOPE</td>
<td>0.620(0.634)</td>
<td>0.859(0.822)</td>
<td>0.749(0.878)</td>
</tr>
<tr>
<td></td>
<td>Kallus and Zhou (2018)</td>
<td>1.061(1.124)</td>
<td>1.363(1.131)</td>
<td>1.679(1.032)</td>
</tr>
<tr>
<td></td>
<td>Colangelo and Lee (2020)</td>
<td>1.827(1.371)</td>
<td>2.292(1.458)</td>
<td>2.429(1.541)</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Deep Jump Learning</td>
<td>2.014(0.865)</td>
<td>1.410(0.987)</td>
<td>1.184(0.967)</td>
</tr>
<tr>
<td>V = 4.86</td>
<td>SLOPE</td>
<td>3.660(0.496)</td>
<td>3.185(0.592)</td>
<td>2.897(0.781)</td>
</tr>
<tr>
<td></td>
<td>Kallus and Zhou (2018)</td>
<td>2.196(2.369)</td>
<td>2.758(2.510)</td>
<td>3.573(2.862)</td>
</tr>
<tr>
<td></td>
<td>Colangelo and Lee (2020)</td>
<td>2.586(2.825)</td>
<td>3.172(3.027)</td>
<td>3.949(3.391)</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Deep Jump Learning</td>
<td>0.494(0.485)</td>
<td>0.412(0.426)</td>
<td>0.349(0.383)</td>
</tr>
<tr>
<td>V = 1.60</td>
<td>SLOPE</td>
<td>0.586(0.337)</td>
<td>0.537(0.279)</td>
<td>0.483(0.272)</td>
</tr>
<tr>
<td></td>
<td>Kallus and Zhou (2018)</td>
<td>2.192(1.210)</td>
<td>2.740(1.034)</td>
<td>3.354(1.324)</td>
</tr>
<tr>
<td></td>
<td>Colangelo and Lee (2020)</td>
<td>2.975(1.789)</td>
<td>3.282(1.525)</td>
<td>3.921(1.927)</td>
</tr>
</tbody>
</table>
Figure B.1: The absolute error of the estimated value and the computational cost (in minutes) under the DJL with different initial number of intervals ($m$) when $n = 100$ in Scenario 1.

Table B.4: The averaged size of the final estimated partition ($|\hat{D}|$) in comparison to the initial number of intervals ($m$) under the proposed DJL for Scenario 1 to 4.

| $|\hat{D}|/m$ | $n=50$ | $n=100$ | $n=200$ | $n=300$ |
|-----------|--------|--------|--------|--------|
| Scenario 1 | 3/5    | 4/10   | 6/20   | 6/30   |
| Scenario 2 | 4/5    | 6/10   | 9/20   | 11/30  |
| Scenario 3 | 4/5    | 6/10   | 8/20   | 10/30  |
| Scenario 4 | 4/5    | 6/10   | 8/20   | 10/30  |

$^1MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2$. See https://en.wikipedia.org/wiki/Mean_squared_error.

$NRMSD = \frac{\sqrt{MSE}}{\max(Y) - \min(Y)}$. See https://en.wikipedia.org/wiki/Root-mean-square_deviation.

$MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}_i|$. See https://en.wikipedia.org/wiki/Mean_absolute_error.

$NMAE = \frac{\max(Y) - \min(Y)}{MAE}$. See https://en.wikipedia.org/wiki/Root-mean-square_deviation.
Table B.5: The mean squared error (MSE), the normalized root-mean-square-deviation (NRMSD), the mean absolute error (MAE), and the normalized MAE (NMAE) of the fitted model under the multilayer perceptrons regressor, linear regression, and the random forest algorithm, via ten-fold cross-validation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Multilayer Perceptrons Regressor</th>
<th>Linear Regression</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.06</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>NRMSD</td>
<td>0.13</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>MAE</td>
<td>0.19</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>NMAE</td>
<td>0.10</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

B.3 Rate of Convergence of Kernel-Based Estimators

B.3.1 Convergence Rate under Model 1

Consider the following piecewise constant function $Q$

$$Q(x, a) = \begin{cases} 
0, & \text{if } a \leq 1/2, \\
1, & \text{otherwise.} 
\end{cases}$$

Define a policy $d$ such that the density function of $d(X)$ equals

$$d(X) = \begin{cases} 
4/3, & \text{if } 1/4 \leq d(X) \leq 1/2, \\
2/3, & \text{else if } 1/2 \leq d(X) < 4/3, \\
0, & \text{otherwise.} 
\end{cases}$$

We aim to show for such $Q$ and $d$, the best possible convergence rate of kernel-based estimator is $n^{-1/3}$.

We first consider its variance. Suppose the conditional variance of $Y|A, X$ is uniformly bounded away from 0. Similar to Theorem 1 of Colangelo and Lee (2020), we can show the variance of kernel based estimator is lower bounded by $O(1)(nh)^{-1}$ where $O(1)$ denotes some positive constant.

We next consider its bias. Since the behavior policy is known, the bias is equal to

$$E\left( \frac{K[[A-d(X)]/h]}{hp(a|x)}[Y-Q(x, d(X))] \right) = E\left( \frac{K[[A-d(X)]/h]}{hp(a|x)}[Q(X, A)-Q(x, d(X))] \right)$$

$$= E\left( \int_{d(x)-h/2}^{d(x)+h/2} K \left( \frac{a-d(X)}{h} \right) \{I[d(X) \leq 1/2 < a] - I[a \leq 1/2 < d(X)] \} da \right).$$
Using the change of variable $a = h t + d(X)$, the bias equals

$$
E\left( \int_{-1/2}^{1/2} K(t)[I(d(X) \leq 1/2 < d(X) + h t) - I(d(X) + h t \leq 1/2 < d(X))] dt \right).
$$

Consider any $0 < h \leq \epsilon$ for some sufficiently small $\epsilon > 0$. The bias is then equal to

$$
\frac{4}{3} \int_{1/2-\epsilon/2}^{1/2} \int_{-1/2}^{1/2} K(t)[I(a \leq 1/2 < a + h t) - I(a + h t \leq 1/2 < a)] dt \, da
$$

$$
+ \frac{2}{3} \int_{1/2}^{1/2+\epsilon/2} \int_{-1/2}^{1/2} K(t)[I(a \leq 1/2 < a + h t) - I(a + h t \leq 1/2 < a)] dt \, da.
$$

Under the symmetric condition on the kernel function, the above quantity is equal to

$$
\frac{2}{3} \int_{1/2-\epsilon/2}^{1/2} \int_{(1-2a)/2h}^{1/2} K(t) dt \, da \geq \frac{2}{3} \int_{1/2-h/4}^{1/2} \int_{(1-2a)/2h}^{1/2} K(t) dt \, da
$$

$$
\geq \frac{2}{3} \int_{1/2-h/2}^{1/2} \int_{1/4}^{1/2} K(t) dt \, da = \frac{h}{6} \int_{1/4}^{1/2} K(t) dt.
$$

Consequently, the bias is lower bounded by $O(1)h$ where $O(1)$ denotes some positive constant.

To summarize, the root mean squared error of kernel based estimator is lower bounded by $O(1)((nh)^{-1/2} + h)$ where $O(1)$ denotes some positive constant. The optimal choice of $h$ that minimizes such lower bound would be of the order $n^{-1/3}$. Consequently, the convergence rate is lower bounded by $O(1)n^{-1/3}$.

### B.3.2 Convergence Rate under Model 2

Similar to the case under Model 1, we can show the variance of kernel-based estimator is lower bounded by $O(n^{-1}h^{-1})$ in cases where the conditional variance of $Y$ given $(A, X)$ is uniformly bounded away from zero.

Consider the conditional mean function $Q$

$$
Q(x, a) = Ch^{-1}K\left\{\frac{a - d(x)}{h}\right\},
$$

for some constant $C > 0$. We aim to derive the bias of kernel-based estimator under such a choice of the conditional mean function $Q$. Using similar arguments in the case where
Model 1 holds, we can show the bias equals
\[ E \left( C^{-1} \frac{K^2 [A - d(X)] / h}{h^2 p(a|x)} \right) \geq C^{-1} E \left( \frac{K^2 [A - d(X)] / h}{h^2} \right). \]

Similarly, we can show the right-hand-side is lower bounded by \( O(1)h \). This implies that the convergence rate is at least \( O(1)(n^{-1} h^{-1} + h) \) under Model 2.

### B.4 Technical Proofs

Throughout the proof, we use \( c, C, c_0, \bar{c}, c_* \), etc., to denote some universal constants whose values are allowed to change from place to place. Let \( O_i = \{X_i, Y_i\} \) denote the data summarized from the \( i \)-th observation. For any two positive sequences \( \{a_n\}_n \) and \( \{b_n\}_n \). The notation \( a_n \sim b_n \) means that there exists some universal constant \( c > 1 \) such that \( c^{-1} b_n \leq a_n \leq c b_n \) for any \( n \). The notation \( a_n \preceq b_n \) means that there exists some universal constant \( c > 0 \) such that \( a_n \leq c b_n \) for all \( n \).

Proofs of Theorems 3.4.1 and 3.4.2 rely on Lemmas B.4.1, B.4.2 and B.4.3. In particular, Lemma B.4.1 establishes the uniform convergence rate of \( \hat{q}_{\mathcal{I}}^{(\ell)} \) for any \( \mathcal{I} \) whose length is no shorter than \( o(\gamma_n) \) and belongs to the set of intervals:

\[
\mathcal{I}(m) = \{ [i_1/m, i_2/m] : \text{for some integers } i_1 \text{ and } i_2 \text{ that satisfy } 0 \leq i_1 < i_2 < m \} \\
\quad \cup \quad \{ [i_3/m, 1] : \text{for some integers } i_3 \text{ that satisfy } 0 \leq i_3 < m \}.
\]

To state this lemma, we first introduce some notations. For any such interval \( \mathcal{I} \), define the function \( q_{\mathcal{I},0}(x) = E(Y|A \in \mathcal{I}, X = x) \). It is immediate to see that the definition of \( q_{\mathcal{I},0} \) here is consistent with the one defined in (3.4) for any \( \mathcal{I} \subseteq \mathcal{D}_0 \).

**Lemma B.4.1.** Assume either conditions in Theorem 1 or 2 are satisfied. Then there exists some constant \( \bar{C} > 0 \) such that the following holds with probability at least \( 1 - O(n^{-2}) \): For any \( 1 \leq \ell \leq \mathcal{L} \), \( \mathcal{I} \in \mathcal{I}(m) \) and \( |\mathcal{I}| \geq c \gamma_n \),

\[ E[|q_{\mathcal{I},0}(X) - \hat{q}_{\mathcal{I}}^{(\ell)}(X)|^2 \{O_i\}_{i \in \mathcal{L}_i}] \leq \bar{C}(n|\mathcal{I}|)^{-2\beta/(2\beta + p)} \log^8 n. \]  \( \text{(B.1)} \)

Here, the expectation in (B.1) is taken with respect to a testing sample \( X \).

**Lemma B.4.2.** Assume either conditions in Theorem 1 or 2 are satisfied. Then there exists some constant \( \bar{C} > 0 \) such that the followings hold with probability at least \( 1 - O(n^{-2}) \): For
any $1 \leq \ell \leq \mathcal{L}$, $\mathcal{I} \in \mathcal{I}(m)$ and $|\mathcal{I}| \geq c \gamma_n$, 

\[
\sum_{\mathcal{I} \in \mathcal{I}(m)} \sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I},0}(X_i)| \{\tilde{q}_{\mathcal{I}}^{(\ell)}(X_i) - q_{\mathcal{I},0}(X_i)\} \leq \tilde{C}(n, |\mathcal{I}|)^{p/(2\beta + p)} \log^6 n.
\]

**Lemma B.4.3.** Assume either conditions in Theorem 1 or 2 are satisfied. Then the following events occur with probability at least $1 - O(n^{-2})$: there exists some constant $c > 0$ such that $\min_{\mathcal{I} \in \mathcal{I}(m)} |\mathcal{I}| \geq c \gamma_n$ for any $1 \leq \ell \leq \mathcal{L}$.

We first present the proofs for these three lemmas. Next we present the proofs for Theorems 3.4.1 and 3.4.2.

**B.4.1 Proof of Lemma B.4.1**

The number of folds $\mathcal{L}$ is bounded. It suffices to derive the uniform convergence rate for each $\ell$. By definition, $\tilde{q}_{\mathcal{I}}^{(\ell)}$ is the minimizer of the least square loss, $\arg\min_{q \in \mathcal{Q}} \sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q(X_i)|^2$. It follows that

\[
\sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|Y_i - \tilde{q}_{\mathcal{I}}^{(\ell)}(X_i)|^2 \leq \sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q(X_i)|^2,
\]

for all $q \in \mathcal{Q}$. Recall that $q_{\mathcal{I},0}(x) = \mathbb{E}[Y | A \in \mathcal{I}, X = x]$, we have $\mathbb{E}[\mathbb{I}(A \in \mathcal{I})|Y - q_{\mathcal{I},0}(X)| | X] = 0$. A simple calculation yields

\[
\sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|q_{\mathcal{I},0}(X_i) - \tilde{q}_{\mathcal{I}}^{(\ell)}(X_i)|^2 \leq \sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|q_{\mathcal{I},0}(X_i) - q(X_i)|^2
\]

\[
+ 2 \sum_{i \in \mathcal{L}_i} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I},0}(X_i)| \{\tilde{q}_{\mathcal{I}}^{(\ell)}(X_i) - q_{\mathcal{I},0}(X_i)\},
\]

for any $q$ and $\mathcal{I}$.

The first term on the right-hand-side measures the approximation bias of the class of deep neural networks. Since $\mathbb{E}[\mathbb{I}(A \in \mathcal{I})|Y - q_{\mathcal{I},0}(X)| | X] = 0$, the second term corresponds to the stochastic error. The rest of the proof is divided into three parts. In Part 1, we bound the approximation error. In Part 2, we bound the stochastic error. Finally, we combine these two parts together to derive the uniform convergence rate for $\tilde{q}_{\mathcal{I}}^{(\ell)}$.

**Part 1.** Under the given condition, we have $Q(\bullet, a) \in \Phi(\beta, c)$, $p(a \bullet) \in \Phi(\beta, c)$ for some $c > 0$ and any $a$. We now argue that there exists some constant $C > 0$ such that $q_{\mathcal{I},0} \in \Phi(\beta, C)$ for
any \( \mathcal{J} \). This can be proven based on the relation that

\[
q_{\mathcal{J},0}(x) = \frac{\int_{\mathcal{J}} Q(x,a) p(a|x) \, da}{\int_{\mathcal{J}} p(a|x) \, da}.
\]

Specifically, we have that \( \sup_x |q_{\mathcal{J},0}(x)| \leq \sup_{a,x} |Q(x,a)| \leq c \). Suppose \( \beta \leq 1 \). For any \( x_1, x_2 \in \mathcal{X} \), consider the difference \( |q_{\mathcal{J},0}(x_1) - q_{\mathcal{J},0}(x_2)| \). Under the positivity assumption, we have \( \inf_{a,x} p(a|x) \geq c_* \) for some \( c_* > 0 \). It follows that

\[
|q_{\mathcal{J},0}(x_1) - q_{\mathcal{J},0}(x_2)| \leq \frac{\int_{\mathcal{J}} |Q(x_1,a) - Q(x_2,a)| p(a|x_1) \, da}{\int_{\mathcal{J}} p(a|x_1) \, da}
+ \frac{\int_{\mathcal{J}} |Q(x_2,a)| p(a|x_1) - p(a|x_2) | p(a|x_2) \, da}{\int_{\mathcal{J}} p(a|x_1) \, da} + \frac{\int_{\mathcal{J}} |Q(x_2,a)| p(a|x_2) \, da}{\int_{\mathcal{J}} p(a|x_2) \, da}
\]

\[
\leq c ||x_1 - x_2||^{\beta - \lfloor \beta \rfloor} + 2 \frac{c^2}{c_*} ||x_1 - x_2||^{\beta - \lfloor \beta \rfloor}.
\]

Consequently, \( q_{\mathcal{J},0} \in \Phi(\beta, c + 2c^2/c_*^2) \).

Suppose \( \beta > 1 \). Then both \( Q(\bullet, a) \) and \( p(a|\bullet) \) are \( |\beta| \)-differentiable. By changing the order of integration and differentiation, we can show that \( q_{\mathcal{J},0}(x) \) is \( |\beta| \)-differentiable as well. As an illustration, when \( \beta < 2 \), we have \( |\beta| = 1 \). According to the chain rule, we have

\[
\frac{\partial q_{\mathcal{J},0}(x)}{\partial x^j} = \frac{\int_{\mathcal{J}} \partial Q(x,a)/\partial x^j | p(a|x) \, da}{\int_{\mathcal{J}} p(a|x) \, da} + \frac{\int_{\mathcal{J}} Q(a|x) | \partial p(a|x)/\partial x^j | \, da}{\int_{\mathcal{J}} p(a|x) \, da}
\]

Moreover, using similar arguments in proving \( q_{\mathcal{J},0} \in \Phi(\beta, c + 2c^2/c_*^2) \) when \( \beta < 1 \), we can show that all the partial derivatives of \( q_{\mathcal{J},0}(x) \) up to the \( |\beta| \)th order are uniformly bounded for all \( \mathcal{J} \). In addition, all the \( |\beta| \)th order partial derivatives are Hölder continuous with exponent \( \beta - |\beta| \). This implies that \( q_{\mathcal{J},0} \in \Phi(\beta, C) \) for some constant \( C > 0 \) and any \( \mathcal{J} \).

It is shown in Lemma 7 of Farrell et al. (2021) that for any \( \epsilon > 0 \), there exists a deep neural network architecture that approximates \( q_{\mathcal{J},0} \) with the uniform approximation error upper bounded by \( \epsilon \), and satisfies \( W_{\mathcal{J}} \leq \hat{C} \epsilon^{-p/\beta} (\log \epsilon^{-1} + 1) \) and \( L_{\mathcal{J}} \leq \hat{C} (\log \epsilon^{-1} + 1) \) for some constant \( \hat{C} > 0 \). These upper bounds will be used later in Part 2. The detailed value of \( \epsilon \) will
be specified below. It follows that for any $\mathcal{S}$, the bias term can be upper bounded by

$$
\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - q(X_i)|^2 \leq \epsilon^2 \sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}).
$$

We next provide an upper bound for the right-hand-side. Since $A$ has a bounded probability density function, the variance $\text{Var}(\mathbb{I}(A_i \in \mathcal{S}))$ is upper bounded by $\sqrt{\mathbb{E}(A_i \in \mathcal{S})} \leq \tilde{c} \sqrt{|\mathcal{S}|}$ for some universal constant $\tilde{c} > 0$. It follows from Bernstein's inequality that

$$
\Pr \left\{ \sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) - |L_c^\mathcal{S}| \mathbb{E}(A \in \mathcal{S}) \geq t \right\} \leq \exp \left( -\frac{t^2/2}{\tilde{c}^2|L_c^\mathcal{S}||\mathcal{S}| + t/3} \right),
$$

for any $t$ and $\mathcal{S}$. Set $t_\mathcal{S} = 6 \max(\tilde{c} \sqrt{n|\mathcal{S}| \log n}, |\mathcal{S}| \log n)$, the right-hand-side is upper bounded by $n^{-4}$. Since $m \simeq n$ and the number of intervals $\mathcal{S}$ in $\mathcal{J}(m)$ is upper bounded by $m^2$, it follows from Bonferroni's inequality that

$$
\Pr \left[ \bigcup_{\mathcal{S} \in \mathcal{J}(m)} \left\{ \sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) - |L_c^\mathcal{S}| \mathbb{E}(A \in \mathcal{S}) \geq t_\mathcal{S} \right\} \right] \leq m^2 n^{-4} = O(n^{-2}).
$$

As such, with probability at least $1 - O(n^{-2})$, we have that $\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) - |L_c^\mathcal{S}| \mathbb{E}(A \in \mathcal{S}) \leq t_\mathcal{S}$ uniformly for all $\mathcal{S}$, or equivalently, $\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) \leq |L_c^\mathcal{S}| \tilde{c} |\mathcal{S}| + t_\mathcal{S}$. Consider a subset of intervals $\mathcal{S}$ with $|\mathcal{S}| \geq c \gamma_n$ for any constant $c > 0$. Under the given conditions on $\gamma_n$, we have

$$
\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) \leq n \tilde{c}^* |\mathcal{S}|, \quad \text{for any } \mathcal{S} \text{ such that } |\mathcal{S}| \geq c \gamma_n,
$$

for some constant $\tilde{c}^* > 0$. It follows from (B.2) that the following holds with probability at least $1 - O(n^{-2})$: for any $\mathcal{S} \in \mathcal{J}(m)$ such that $|\mathcal{S}| \geq c \gamma_n$, we have

$$
\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - q(X_i)|^2 \leq \tilde{c}^* n |\mathcal{S}|.
$$

Set $\epsilon$ to $(n|\mathcal{S}|)^{-\beta/(2\beta+p)}$, it follows that

$$
\sum_{i \in L_c^\mathcal{S}} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - q(X_i)|^2 \leq \epsilon^*(n|\mathcal{S}|)^{-2\beta/(2\beta+p)} (n|\mathcal{S}|).
$$

(B.4)
$W_\mathcal{F}$ and $L_\mathcal{F}$ are upper bounded by $\tilde{C}(n|\mathcal{F}|)^p/(2\beta+p)(\beta \log(n|\mathcal{F}|)/(2\beta+p)+1)$ and $\tilde{C}(\beta \log(n|\mathcal{F}|)/(2\beta+p)+1)$, respectively. This completes the proof for Part 1.

**Part 2.** For the function class of deep neural networks $Q_\mathcal{F}$, we use $\theta_\mathcal{F}$ to denote the parameters in deep neural networks. This allows us to represent $\mathcal{Q}_\mathcal{F}$ as $\{q_\mathcal{F}(\bullet, \theta_\mathcal{F}) : \theta_\mathcal{F}\}$. We will apply the empirical process theory (see e.g., Van Der Vaart and Wellner 1996) to bound the stochastic error. Let $\hat{\theta}_\mathcal{F}$ be the estimated parameter in $\mathcal{Q}_\mathcal{F}^{(l)}$. Define

$$\sigma^2(\mathcal{F}, \theta) = E\{\mathbb{I}(A \in \mathcal{F})|q_{\mathcal{F},0}(X) - q_{\mathcal{F},0}(X, \theta)|^2\},$$

for any $\theta$ and $\mathcal{F}$. Consider two separate cases, corresponding to $\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F}) \leq |\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}$ and $\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F}) > |\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}$, respectively. We focus our attentions on the latter class of intervals. In Part 3, we will show that for those intervals,

$$\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F}) \leq O(1)|\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)} \log^4 n,$$

for some universal constant $O(1)$. This implies that for any $\mathcal{F}$, we have

$$\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F}) \leq O(1)|\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)} \log^4 n. \tag{B.5}$$

We consider bounding a scaled version of the stochastic error,

$$\frac{1}{\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F})} \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{F})\{Y_i - q_{\mathcal{F},0}(X_i)\}\{q_{\mathcal{F},0}(X_i, \hat{\theta}_\mathcal{F}) - q_{\mathcal{F},0}(X_i)\}. $$

Its absolute value can be upper bounded by

$$Z(\mathcal{F}) \equiv \sup_{\theta} \left| \frac{1}{\sigma(\mathcal{F}, \theta)} \sum_{i \in \mathcal{I}} \mathbb{I}(A_i \in \mathcal{F})\{Y_i - q_{\mathcal{F},0}(X_i)\}\{q_{\mathcal{F},0}(X_i, \theta) - q_{\mathcal{F},0}(X_i)\} \right|,$$

where the supremum is taken over all $\theta$ such that $\sigma(\mathcal{F}, \theta) > |\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}$.

For a given $\theta$, the empirical sum has zero mean. Under the boundedness assumption on $Y$, its variance is upper bounded by some universal constant. In addition, each quantity $\sigma^{-1}(\mathcal{F}, \theta)\mathbb{I}(A_i \in \mathcal{F})\{Y_i - q_{\mathcal{F},0}(X_i)\}\{q_{\mathcal{F},0}(X_i, \theta) - q_{\mathcal{F},0}(X_i)\}$ is upper bounded by $O(1)|\mathcal{F}|^{-1/2}(n|\mathcal{F}|)^{\beta/(2\beta+p)}$ for some universal constant $O(1)$. This allows us to apply the tail inequality developed by Massart et al. (2000) to bounded the empirical process. See also Theorem 2 of Adamczak et al. (2008). Specifically, for all $t > 0$ and $\mathcal{F}$ that satisfies $\sigma(\mathcal{F}, \hat{\theta}_\mathcal{F}) > |\mathcal{F}|^{1/2}(n|\mathcal{F}|)^{-\beta/(2\beta+p)}$, $\ldots$
we obtain with probability at least $1 - \exp(t)$ that

$$Z(I) \leq 2EZ(I) + \tilde{c} \sqrt{t} n + t \tilde{c} |\mathcal{I}|^{-1/2}(n|\mathcal{I}|)^{\beta/(2\beta + p)},$$

(B.6)

for some constant $\tilde{c} > 0$. By setting $t = 3 \log n$, the probability $1 - \exp(t) = 1 - n^{-3}$. Notice that the number of intervals $I$ is upper bounded by $O(n^2)$, under the condition that $m$ is proportional to $n$. By Bonferroni’s inequality, we obtain that (B.6) holds with probability at least $1 - O(n^{-2})$ for any $I$. Under the given condition on $\gamma_n$, for any interval $I$ such that $|\mathcal{I}| \geq c \gamma_n$, the last term on the right-hand-side of (B.6) is $o(\sqrt{n})$. It follows that the following occurs with probability $1 - O(n^{-2})$,

$$Z(I) \leq 2EZ(I) + 2 \tilde{c} \sqrt{n \log n},$$

(B.7)

for all $I$ such that $|\mathcal{I}| \geq c \gamma_n$ and $\sigma(I, \hat{\theta}_I) > |\mathcal{I}|^{-1/2}(n|\mathcal{I}|)^{-\beta/(2\beta + p)}$.

We next provide an upper bound for $EZ(I)$. Toward that end, we will apply the maximal inequality developed in Corollary 5.1 of Chernozhukov et al. (2014). We first observe that the class of empirical sum indexed by $\theta$ belongs to the VC subgraph class with VC-index upper bounded by $O(W L \log(W \log n))$. It follows that for any $I$ such that $|\mathcal{I}| \geq c \gamma_n$, $\sigma(I, \hat{\theta}_I) > |\mathcal{I}|^{-1/2}(n|\mathcal{I}|)^{-\beta/(2\beta + p)}$,

$$EZ(I) \propto \sqrt{n \log n \log(W \log n)} + W \log(W \log n).$$

Based on the upper bounds on $W$ and $L$ developed in Part 1, the right-hand-side is upper bounded by

$$O(1)(n|\mathcal{I}|^{p/(4\beta + 2p)} \sqrt{n \log^4 n} + O(1)|\mathcal{I}|^{-1/2}(n|\mathcal{I}|)^{p/(2\beta + p)} \log^4 n),$$

where $O(1)$ denotes some universal constant. It is of the order $O\{n^{1/2}(n|\mathcal{I}|)^{p/(4\beta + 2p)} \log^4 n\}$. This yields that

$$EZ(I) \propto n^{1/2}(n|\mathcal{I}|)^{p/(4\beta + 2p)} \log^4 n.$$

This together with (B.6) and (B.7) yields that with probability at least $1 - O(n^{-2})$, the scaled stochastic error is upper bounded by $n^{1/2}(n|\mathcal{I}|)^{p/(4\beta + 2p)} \log^4 n$. As such, with probability at
least $1 - O(n^{-2})$, we obtain that

$$\left| \sum_{i \in \mathcal{L}_1^c} \mathbb{I}(A_i \in \mathcal{S}) [Y_i - q_{\mathcal{S},0}(X_i)] \{ \hat{q}^{(l)}_{\mathcal{S}}(X_i) - q_{\mathcal{S},0}(X_i) \} \right| \leq n \sigma^2(\mathcal{S}, \hat{\theta}_s) n^{\frac{1}{2}} (n|\mathcal{S}|)^{\frac{p}{(2\beta + p)}} \log^4 n,$$

for any $\mathcal{S}$ such that $|\mathcal{S}| \geq c \gamma_n$, $\sigma(\mathcal{S}, \hat{\theta}_s) > |\mathcal{S}|^{1/2} (n|\mathcal{S}|)^{-\beta/(2\beta + p)}$. By Cauchy-Schwarz inequality, the left-hand-side can be further upper bounded by

$$\frac{n \sigma^2(\mathcal{S}, \hat{\theta}_s)}{4} + O(1) (n|\mathcal{S}|)^{p/(2\beta + p)} \log^8 n,$$

where $O(1)$ denotes some universal positive constant. This completes the proof for Part 2.

**Part 3.** Combining the results in Part 1 and Part 2, we obtain that for any $\mathcal{S}$ such that $|\mathcal{S}| \geq c \gamma_n$, $\sigma(\mathcal{S}, \hat{\theta}_s) > |\mathcal{S}|^{1/2} (n|\mathcal{S}|)^{-\beta/(2\beta + p)}$,

$$\sum_{i \in \mathcal{L}_1^c} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - \hat{q}^{(l)}_{\mathcal{S}}(X_i)|^2 \leq \frac{n \sigma^2(\mathcal{S}, \hat{\theta}_s)}{4} + O(1) (n|\mathcal{S}|)^{p/(2\beta + p)} \log^8 n,$$

with probability at least $1 - O(n^{-2})$. As for the left-hand-side, we notice that

$$\sum_{i \in \mathcal{L}_1^c} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - \hat{q}^{(l)}_{\mathcal{S}}(X_i)|^2 \geq |\mathcal{L}_1^c | \sigma^2(\mathcal{S}, \hat{\theta}_s) - \left| \sum_{i \in \mathcal{L}_1^c} \mathbb{I}(A_i \in \mathcal{S}) |q_{\mathcal{S},0}(X_i) - \hat{q}^{(l)}_{\mathcal{S}}(X_i)|^2 - |\mathcal{L}_1^c | \sigma^2(\mathcal{S}, \hat{\theta}_s) \right|.$$

Using similar arguments in Part 2, we can show that the second line is upper bounded by $n \sigma^2(\mathcal{S}, \hat{\theta}_s)/8 + O(1) (n|\mathcal{S}|)^{p/(2\beta + p)} \log^8 n$, with probability at least $1 - O(n^{-2})$, for any $\mathcal{S}$ such that $|\mathcal{S}| \geq c \gamma_n$, $\sigma(\mathcal{S}, \hat{\theta}_s) > |\mathcal{S}|^{1/2} (n|\mathcal{S}|)^{-\beta/(2\beta + p)}$. Since $\mathcal{L}_1^c \geq n/2$, we obtain

$$\left( \frac{1}{2} - \frac{1}{4} - \frac{1}{8} \right) \sigma^2(\mathcal{S}, \hat{\theta}_s) = \frac{1}{8} \sigma^2(\mathcal{S}, \hat{\theta}_s) \propto (n|\mathcal{S}|)^{-2\beta/(2\beta + p)} \log^8 n.$$

This yields the desired uniform upper bound for $\sigma^2(\mathcal{S}, \hat{\theta}_s)$. We thus obtain (B.5) holds with probability at least $1 - O(n^{-2})$.

Under the assumption that the density function $p(a|x)$ is uniformly bounded away from zero, we obtain

$$\sigma^2(\mathcal{S}, \hat{\theta}_s) \leq c |\mathcal{S}| E|q_{\mathcal{S},0}(X) - \hat{q}^{(l)}_{\mathcal{S}}(X)|^2,$$
for some constant $c > 0$. This assertion thus follows.

### B.4.2 Proof of Lemma B.4.2

The assertion can be proven in a similar manner as Part 2 of the proof of Lemma B.4.1. We omit the details to save space.

### B.4.3 Proof of Lemma B.4.3

Consider a given interval $\mathcal{I} \in \mathcal{I}^{(l)}$. Suppose $|\mathcal{I}| < c \gamma_n$. The value of the constant $c$ will be determined later. Then, for sufficiently large $n$, we can find some interval $\mathcal{I}' \in \mathcal{I}(m) \cap \mathcal{I}^{(l)}$ that is adjacent to $\mathcal{I}$. Thus, we have $\mathcal{I} \cup \mathcal{I}' \in \mathcal{I}(m)$, and hence

\[
\frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - \hat{q}_{\mathcal{I}, \mathcal{I}'}^{(l)}(X_i) \}^2 + \frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I}') \{ Y_i - \hat{q}_{\mathcal{I}, \mathcal{I}'}^{(l)}(X_i) \}^2 \leq \frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') \{ Y_i - \hat{q}_{\mathcal{I} \cup \mathcal{I}'}^{(l)}(X_i) \}^2 - \gamma_n. \tag{B.8}
\]

Notice that the left-hand-side of the above expression is nonnegative. It follows that

\[
\gamma_n \leq \frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') \{ Y_i - \hat{q}_{\mathcal{I} \cup \mathcal{I}'}^{(l)}(X_i) \}^2.
\]

By definition, we have

\[
\hat{q}_{\mathcal{I} \cup \mathcal{I}'}^{(l)} = \arg \min_{q_{\mathcal{I} \cup \mathcal{I}'} \in \mathcal{Q}} \frac{1}{n} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') \{ Y_i - q_{\mathcal{I} \cup \mathcal{I}'}(X_i) \}^2.
\]

It follows that

\[
\frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') \{ Y_i - \hat{q}_{\mathcal{I} \cup \mathcal{I}'}^{(l)}(X_i) \}^2 \leq \frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I} \cup \mathcal{I}') \{ Y_i - \hat{q}_{\mathcal{I} \cup \mathcal{I}'}^{(l)}(X_i) \}^2.
\]

By (B.8), this further implies that

\[
\frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - \hat{q}_{\mathcal{I}}^{(l)}(X_i) \}^2 \leq \frac{1}{|L^c_{\ell}|} \sum_{i \in L^c_{\ell}} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - \hat{q}_{\mathcal{I}}^{(l)}(X_i) \}^2 - \gamma_n.
\]
and hence
\[ \gamma_n \leq \frac{1}{|L^c|} \sum_{i \in L^c} \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - \tilde{q}^{(\ell)}_{\mathcal{S}}(X_i) \}^2. \]

Under (A2), the function \( \tilde{q}_{\mathcal{S}} \) is uniformly upper bounded from above. It thus follows from Cauchy-Schwarz inequality that
\[ \gamma_n \leq \frac{2}{|L^c|} \sum_{i \in L^c} \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i^2 + \tilde{q}^{2}_{\mathcal{S}}(X_i) \} \leq c_0 n^{-1} \sum_{i \in L^c} \mathbb{I}(A_i \in \mathcal{S}), \]
for some constant \( c_0 > 0 \). Using similar arguments in showing (B.3), we can show that with probability at least \( 1 - O(n^{-2}) \), the following even holds for all \( \mathcal{S} \in \mathcal{I}(m) \),
\[ n^{-1} \sum_{i \in L^c} \mathbb{I}(A_i \in \mathcal{S}) \leq c_1 (\sqrt{n^{-1}} |\mathcal{S}| \log n + |\mathcal{S}|), \]
for some constant \( c_1 > 0 \). The right-hand-side shall be larger than or equal to \( \gamma_n \). Consequently, we have either \( |\mathcal{S}| \geq c_2 \gamma_n \) or \( |\mathcal{S}| \geq c_2 n \gamma^2_n / \log n \) for some constant \( c_2 > 0 \). Under the given condition on \( \gamma_n \), we obtain that \( |\mathcal{S}| \geq c_2 \gamma_n \) for sufficiently large \( n \). The proof is hence completed.

**B.4.4 Proof of Theorem 3.4.1**

Since the number of folds \( L \) is a fixed integer. We will show the assertions in (i) and (ii) holds for each \( \ell \), with probability at least \( 1 - O(n^{-2}) \). The proof is divided into three parts. In Part 1, we show the consistency of the estimated change point locations and that \( |\hat{\mathcal{S}}^{(\ell)}| \geq |\mathcal{S}_0| \) with probability at least \( 1 - O(n^{-2}) \). In Part 2, we prove that \( |\hat{\mathcal{S}}^{(\ell)}| = |\mathcal{S}_0| \) with probability at least \( 1 - O(n^{-2}) \) and derive the rate of convergence of the estimated change point locations and the estimated function \( Q \). In Part 3, we derive the rate of convergence for the value estimator.

**Part 1.** We first show the consistency of the estimated change-point locations. Assume \( |\mathcal{S}_0| > 1 \). Otherwise, the assertion \( |\hat{\mathcal{S}}^{(\ell)}| \geq |\mathcal{S}_0| \) trivially hold. Consider the partition \( \mathcal{S} = \{ [0, 1] \} \) which consists of a single interval and a zero function \( Q \). By definition, we have
\[ \sum_{\mathcal{S} \in \hat{\mathcal{S}}^{(\ell)}} \left( \sum_{i \in L^c} \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - \tilde{q}^{(\ell)}_{\mathcal{S}}(X_i) \}^2 \right) + |L^c| |\hat{\mathcal{S}}^{(\ell)}| |\mathcal{S}_0| \leq \sum_{i \in L^c} Y_i^2 + |L^c| \gamma_n. \]
Under the boundedness assumption on $Y$, we obtain that $|L_{c}^\ell |\gamma_n|\hat{\Theta}^{(l)}| \leq C_0(|L_{c}^\ell | + \gamma_n)$ for some constant $C_0 > 0$ and hence

$$|\hat{\Theta}^{(l)}| \leq 2C_0\gamma_n^{-1}, \quad \text{(B.9)}$$

for sufficiently large $n$, as $\gamma_n \to 0$.

Notice that

$$
\sum_{\mathcal{S} \in \mathcal{G}(\nu)} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - \hat{q}_\nu^{(l)}(X_i) \}^2 \geq \sum_{\mathcal{S} \in \mathcal{G}(\nu)} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - q_{\mathcal{S},0}(X_i) \}^2 \eta_i^*,
$$

and

$$
-2 \sum_{\mathcal{S} \in \mathcal{G}(\nu)} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - q_{\mathcal{S},0}(X_i) \} \{ \hat{q}_\nu^{(l)}(X_i) - q_{\mathcal{S},0}(X_i) \}.
$$

The second line is non-negative. Under Lemmas B.4.2 and B.4.3, the third line is lower bounded by $-C_1 \sum_{\mathcal{S} \in \mathcal{G}(\nu)} (n |\mathcal{S}|)^{p/(p+2\beta)} \log^8 n$ for some constant $C_1 > 0$ with probability at least $1 - O(n^{-2})$. In view of (B.9), it can be further lower bounded by $-2C_0C_1\gamma_n^{-1} n^{p/(p+2\beta)} \log^8 n$. By (B.9) and the given condition on $\gamma_n$, the third line is $o(n)$. It follows that

$$
\sum_{\mathcal{S} \in \mathcal{G}(\nu)} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - \hat{q}_\nu^{(l)}(X_i) \}^2 \geq \eta_i^* + o(n), \quad \text{(B.10)}
$$

with probability at least $1 - O(n^{-2})$.

Similar to (B.3), we can show that the following events occur with probability at least $1 - O(n^{-2})$,

$$
\left| \frac{1}{|L_{c}^\ell |} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Y_i - Q(X_i, A_i) \} \{ Q(X_i, A_i) - q_{\mathcal{S},0}(X_i) \} \right| \leq c_0 \left[ n^{-1/2} \sqrt{\mathbb{E} \mathbb{I}(A \in \mathcal{S}) |Q(X, A) - q_{\mathcal{S},0}(X)|^2 \log n + n^{-1} \log n \right], \quad \text{(B.11)}
$$

and

$$
\left| \frac{1}{|L_{c}^\ell |} \sum_{i \in L_{c}^\ell } \mathbb{I}(A_i \in \mathcal{S}) \{ Q(X_i, A_i) - q_{\mathcal{S},0}(X_i) \}^2 - \mathbb{E} \mathbb{I}(A \in \mathcal{S}) |Q(X, A) - q_{\mathcal{S},0}(X)|^2 \right| \leq c_0 \left[ n^{-1/2} \sqrt{\mathbb{E} \mathbb{I}(A \in \mathcal{S}) |Q(X, A) - q_{\mathcal{S},0}(X)|^2 \log n + n^{-1} \log n \right], \quad \text{(B.12)}
$$

for some constant $c_0 > 0$. For any interval $\mathcal{S}$, the two upper bounds in (B.11) and (B.12) are
It follows that
\[
\eta_1^* = \sum_{\ell \in \mathcal{L}} \sum_{i \in L_{\ell}^c} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \hat{q}_i^f(X_i))^2 + \sum_{\ell \in \mathcal{L}} \mathbb{I}(A_i \in \mathcal{I}) (Q(X_i, A_i) - q_{\mathcal{I},0}(X_i))^2
\]
\[
+ 2 \sum_{\ell \in \mathcal{L}} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - Q(X_i, A_i)) (Q(X_i, A_i) - q_{\mathcal{I},0}(X_i))
\]
\[
= \sum_{i \in L_{\ell}^c} |Y_i - Q(X_i, A_i)|^2 + |L_{\ell}^c| \sum_{\ell \in \mathcal{L}} \mathbb{E}(A \in \mathcal{I}) |Q(X, A) - q_{\mathcal{I}}(X)|^2 + o(n),
\]
with probability at least \(1 - O(n^{-2})\). It follows from (B.10) that
\[
\sum_{\ell \in \mathcal{L}} \sum_{i \in L_{\ell}^c} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - \hat{q}_i^f(X_i))^2 \geq \sum_{i \in L_{\ell}^c} |Y_i - Q(X_i, A_i)|^2
\]
\[
\quad + |L_{\ell}^c| \sum_{\ell \in \mathcal{L}} \mathbb{E}(A \in \mathcal{I}) |Q(X, A) - q_{\mathcal{I}}(X)|^2 + o(n),
\]
with probability at least \(1 - O(n^{-2})\).

Let us consider \(\eta_2^*\). We observe that
\[
\eta_2^* = \sum_{\ell \in \mathcal{L}} \sum_{i \in L_{\ell}^c} \mathbb{I}(A_i \in \mathcal{I}) |Y_i - q_{\mathcal{I},0}(X_i)|^2.
\]
By the uniform approximation property of deep neural networks, there exists some \(q_{\mathcal{I}}^f \in \mathcal{Q}\) such that
\[
\sum_{i \in L_{\ell}^c} |q_{\mathcal{I},0}(X_i) - q_{\mathcal{I}}^f(X_i)|^2 \propto n(n|\mathcal{I}|)^{-2\beta/(2\beta + p)}.
\]
See Part 1 of the proof of Lemma B.4.1 for details. Similar to (B.3), we can show that the following events occur with probability at least \(1 - O(n^{-2})\),
\[
\frac{1}{|L_{\ell}^c|} \sum_{i \in L_{\ell}^c} \mathbb{I}(A_i \in \mathcal{I}) (Y_i - q_{\mathcal{I}}(X_i)) (q_{\mathcal{I}}(X_i) - q_{\mathcal{I}}^f(X_i)) \leq \frac{c_0 \sqrt{|\mathcal{I}| \log n}}{\sqrt{n}} (n|\mathcal{I}|)^{-\beta/(2\beta + p)},
\]
with probability at least \(1 - O(n^{-2})\).
for some constant $c_0 > 0$ and any $\mathcal{I} \in \mathcal{D}_0$. It follows that

$$
\eta_2^* - \sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^*(X_i)|^2 \geq - \sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|q_{\mathcal{I},0}(X_i) - q_{\mathcal{I}}^*(X_i)|^2
$$

$$
-2 \sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^*(X_i)| \{q_{\mathcal{I}}(X_i) - q_{\mathcal{I}}^*(X_i)\} \geq -\tilde{c} n^{p/(2\beta + p)},
$$

for some constant $\tilde{c} > 0$. This together with (B.13) yields that

$$
\sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^{(l)}(X_i)|^2 \geq \sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^*(X_i)|^2
$$

$$
+ |L_f^c| \sum_{\mathcal{I} \in \mathcal{D}_0} \mathbb{I}(A \in \mathcal{I})|Q(X, A) - q_{\mathcal{I},0}(X)|^2 + o(n) + O\{n^{p/(2\beta + p)}\},
$$

(B.14)

with probability at least $1 - O(n^{-2})$.

Let $K = |\mathcal{D}_0|$. For any integer $k$ such that $1 \leq k \leq K - 1$, let $\tau_{0,k}^*$ be the change point location that satisfies $\tau_{0,k}^* = i/m$ for some integer $i$ and that $|\tau_{0,k} - \tau_{0,k}^*| < m^{-1}$. Denoted by $\mathcal{D}^*$ the oracle partition formed by the change point locations $\{\tau_{0,k}^*\}_{k=1}^{K-1}$. Set $\tau_{0,0}^* = 0$, $\tau_{0,K}^* = 1$ and $q_{\mathcal{I}_{\tau_{0,k-1}^*,\tau_{0,k}^*}}^{(\tau_{0,k-1}^*,\tau_{0,k}^*)} = q_{\mathcal{I}_{\tau_{0,k-1}^*,\tau_{0,k}^*}}^*$ for $1 \leq k \leq K - 1$. Let $\Delta_k = [\tau_{0,k-1}^*, \tau_{0,k}^*] \cap [\tau_{0,k-1}, \tau_{0,k}]^c$ for $1 \leq k \leq K - 1$ and $\Delta_K = [\tau_{0,K-1}^*, 1] \cap [\tau_{0,K-1}, 1]^c$. The length of each interval $\Delta_k$ is at most $m^{-1}$. It follows that

$$
\left( \sum_{\mathcal{I} \in \mathcal{D}^*} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^{\mathcal{D}^*}(X_i)|^2 \right) + \gamma_n |L_f^c| |\mathcal{D}^*| \geq \left( \sum_{\mathcal{I} \in \mathcal{D}_0} \sum_{i \in L_I^*} \mathbb{I}(A_i \in \mathcal{I})|Y_i - q_{\mathcal{I}}^*(X_i)|^2 \right) + \gamma_n |L_f^c| |\mathcal{D}_0|
$$

$$
\leq 2 \sum_{k=1}^{K} \sum_{i \in L_I} \mathbb{I}(A_i \in \Delta_k) \left\{ Y_i^2 + \sup_{\mathcal{I} \subseteq [0,1]} q_{\mathcal{I}}^2(X_i) \right\}.
$$

Since $Y$ is a bounded variable, $q_{\mathcal{I}}^*$ is uniformly bounded for any $\mathcal{I}$. The right-hand-side is upper bounded by $\sum_{k=1}^{K} \sum_{i \in L_I} \mathbb{I}(A_i \in \Delta_k)$. Similar to (B.3), The later is upper bounded by
\[ O(\log n), \text{ with probability at least } 1 - O(n^{-2}). \text{ It follows that} \]

\[
\left( \sum_{I \in \mathcal{I}^n} \left[ \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 \right] + \gamma_n \| L_i \| \mathcal{G} \right) 
\]  \[ - \left( \sum_{I \in \mathcal{I}_0} \left[ \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 \right] + \gamma_n \| L_i \| \mathcal{G}_0 \right) \leq O(\log n), \tag{B.15} \]

with probability at least \( 1 - O(n^{-2}). \) By definition,

\[
\sum_{I \in \mathcal{I}^n} \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 + \gamma_n \| L_i \| \mathcal{G} \leq \sum_{I \in \mathcal{I}^n} \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 + \gamma_n \| L_i \| \mathcal{G}. \tag{B.16} \]

Combining this together with (B.15) yields that

\[
\sum_{I \in \mathcal{I}^n} \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 + \gamma_n \| L_i \| \mathcal{G} \leq \sum_{I \in \mathcal{I}_0} \sum_{i \in I} \mathbb{I}(A_i \in \mathcal{I}) \{ Y_i - q^{a_i}(X_i) \}^2 + \gamma_n \| L_i \| \mathcal{G}_0 + O(\log n). \]

It follows from (B.14) and the condition \( \gamma_n \to 0 \) that

\[
\sum_{I \in \mathcal{I}^n} \mathbb{E}(A \in \mathcal{I}) |Q(X, A) - q_{\mathcal{I}, 0}(X)|^2 = o(1), \tag{B.17} \]

with probability at least \( 1 - O(n^{-2}). \) Under the event defined above, we show that \( \max_{\tau \in \mathcal{I}^n} \min_{\tau \in \mathcal{I}^n} |\hat{\tau} - \tau| \leq \delta \) for any constant \( \delta > 0. \) This yields the consistency of our estimated change point locations.

Specifically, under the condition that \( q_{\mathcal{I}, 0} \neq q_{\mathcal{I}, 0} \) for any adjacent \( \mathcal{I}_1, \mathcal{I}_2 \in \mathcal{I}_0, \) we have \( \mathbb{E}|q_{\mathcal{I}, 1}(X) - q_{\mathcal{I}, 0}(X)|^2 > 0. \) Let \( \delta_0 \) denote the minimum distance between two change point locations. Since the change points are fixed, \( \delta_0 \) is a fixed positive value. For a given \( 0 < \delta < \delta_0, \) suppose \( \max_{\tau \in \mathcal{I}^n} \min_{\tau \in \mathcal{I}^n} |\hat{\tau} - \tau| > \delta. \) Then there exists a change point \( \tau_0 \) and \( I \in \mathcal{I}^n \) such that \( \tau_0 \in \mathcal{I}, |\mathcal{I}| \geq 2\delta \) and that \( \min(|a - \tau_0|, |b - \tau_0|) \geq \delta \) where \( a, b \) correspond to the endpoints of the interval \( \mathcal{I}. \) Under the event defined in (B.17), we have

\[
\mathbb{E}(A \in [a, b]) |Q(X, A) - q_{\mathcal{I}, 0}(X)|^2 = o(1). \tag{B.18} \]

205
Since $\delta_0 > \delta$, the conditional mean function $Q$ is a piecewise function of $A$ in the intervals $[a, \tau_0]$ and $[\tau_0, b]$. The left-hand-side thus equals

$$E\{A \in [\tau_0, b]|q_{[\tau_0, b],0}(X) - q_{\tau_0,0}(X)\}^2 + E\{A \in [a, \tau_0]|q_{[a, \tau_0],0}(X) - q_{\tau_0,0}(X)\}^2.$$ 

The function $q_{\tau_0,0}$ that minimizes the above objective is given by

$$\{E\{A \in [a, b]|X\}\}^{-1}\{E\{A \in [a, \tau_0]|X\} + q_{[\tau_0, b],0}(X)E\{A \in [\tau_0, b]|X\}\}.$$ 

Consequently, the left-hand-side of (B.18) is greater than or equal to

$$E\{E\{A \in [\tau_0, b]|X\}\{E\{A \in [a, \tau_0]|X\}|q_{[\tau_0, b],0}(X) - q_{[\tau_0, b],0}(X)\}^2,$$

which is not to decay to zero since $\min(|a - \tau_0|, |b - \tau_0|) \geq \delta$ and that $q_{\tau_0,0} \neq q_{\tau_0,0}$ for any adjacent $\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{D}_0$. This contradicts (B.18). As such, we obtain that $\max_{\tau \in J(\mathcal{D}_0)} \min_{\tau \in J(\mathcal{D}_0)} |\hat{\tau} - \tau| \leq \delta$ for any sufficiently small $\delta$. This yields the consistency of the estimated change point locations. It also implies that $|\hat{\mathcal{D}}(l)| \geq |\mathcal{D}_0|$ with probability at least $1 - O(n^{-2})$. This completes the proof of Part 1.

**Part 2.** In this part, we show $|\hat{\mathcal{D}}(l)| = |\mathcal{D}_0|$ with probability at least $1 - O(n^{-2})$ and derive the rate of convergence of the estimated change point locations. Similar to (B.14) and (B.15), with a more refined analysis (see Part 1 of the proof), we obtain that

$$\sum_{\mathcal{J} \in \hat{\mathcal{D}}(l)} \sum_{i \in \mathcal{I}_i^\mathcal{J}} \mathbb{I}(A_i \in \mathcal{J})|Y_i - \tilde{q}_\mathcal{J}(X_i)|^2 \geq \sum_{\mathcal{J} \in \mathcal{D}_0} \sum_{i \in \mathcal{I}_i^\mathcal{J}} \mathbb{I}(A_i \in \mathcal{J})|Y_i - q^{*\mathcal{J}}_\mathcal{J}(X_i)|^2$$

$$+ |\mathcal{L}_f^\mathcal{J}| \sum_{\mathcal{J} \in \mathcal{D}_0} \mathbb{E}\{A \in \mathcal{J}|Q(X, A) - q_{\mathcal{J},0}(X)\}^2 - C_1|\mathcal{E}(l)|^{\beta/(2p+\beta)}n^{p/(2p+\beta)}\log^8 n + O(n^{p/(2\beta+p)})$$

$$- 2c_0|\mathcal{L}_f^\mathcal{J}|^{1/2} \sum_{\mathcal{J} \in \mathcal{D}_0} \sqrt{\mathbb{E}\{A \in \mathcal{J}|Q(X, A) - q_{\mathcal{J},0}(X)\}^2} \log n - 2c_0|\mathcal{D}(l)|\log n.$$

with probability at least $1 - O(n^{-2})$. By Cauchy-Schwarz inequality, the third line is lower bounded by

$$- \frac{|\mathcal{L}_f^\mathcal{J}|}{2} \sum_{\mathcal{J} \in \mathcal{D}_0} \mathbb{E}\{A \in \mathcal{J}|Q(X, A) - q_{\mathcal{J},0}(X)\}^2 - 2(c_0 + c_0^2)|\mathcal{D}(l)|\log n.$$
It follows that
\[
\sum_{J \in \mathcal{P}(I)} \sum_{I \subseteq J} \mathbb{I}(A_i \in J)[Y_i - \hat{q}_c(X_i)]^2 \geq \sum_{J \in \mathcal{P}(I)} \sum_{I \subseteq J} \mathbb{I}(A_i \in J)[Y_i - q^{**}_c(X_i)]^2
\]
\[
+ \frac{|L^c|}{2} \sum_{J \in \mathcal{P}(I)} \mathbb{E}(A \in J)Q(X, A) - q_{\mathcal{A}_0}(X)^2 - C_1 |\hat{\mathcal{T}}(I)| \beta/(2p+\beta) n^{p/(p+2\beta)} \log^8 n
\]
\[-2(c_0 + c_0^2) |\hat{\mathcal{T}}(I)| \log n + O(n^{p/(2\beta+p)}).
\]
This together with (B.16) yields that
\[
\frac{|L^c|}{2} \sum_{J \in \mathcal{P}(I)} \mathbb{E}(A \in J)Q(X, A) - q_{\mathcal{A}_0}(X)^2 \leq C_1 |\hat{\mathcal{T}}(I)| \beta/(2p+\beta) n^{p/(p+2\beta)} \log^8 n
\]
\[+ O(n^{p/(2\beta+p)}) + n\gamma_n (|\mathcal{D}_0| - |\hat{\mathcal{T}}(I)|) + 2(c_0 + c_0^2) |\hat{\mathcal{T}}(I)| \log n.
\]
Under the given condition on \(\gamma_n\), we obtain that \(|\hat{\mathcal{T}}(I)| \leq |\mathcal{D}_0|\). Combining this together with \(|\hat{\mathcal{T}}(I)| \geq |\mathcal{D}_0|\), we obtain that \(|\hat{\mathcal{T}}(I)| = |\mathcal{D}_0|\). As such, we have that
\[
\sum_{J \in \mathcal{P}(I)} \mathbb{E}(A \in J)Q(X, A) - q_{\mathcal{A}_0}(X)^2 \propto n^{-2\beta/(p+2\beta)} \log^8 n
\]
Using similar arguments in establishing the consistency of the estimated change point locations, we can show that under the above event, we have that \(\max_{J \in \mathcal{D}_0} \min_{\hat{\tau}: J \in \mathcal{T}(\hat{\mathcal{T}})} |\hat{\tau} - \tau| \propto n^{-2\beta/(p+2\beta)} \log^8 n\). This completes the proof of this part.

Part 3. For any target policy \(d\), we define a random policy \(d_{\mathcal{D}_0}\) according to the partition \(\mathcal{T}(I)\) as follows:
\[
d_{\mathcal{D}_0}(a|x) = \sum_{J \in \mathcal{P}(I)} \mathbb{I}(d(x) \in J, a \in J) \frac{p(a|x)}{b(J|x)},
\]
where \(b(J|x)\) denotes the propensity score function \(\text{Pr}(A \in J|X = x)\). Note that \(\int_0^1 d_{\mathcal{D}_0}(a|x) da = \sum_{J \in \mathcal{D}_0} \mathbb{I}(d(x) \in J) = 1\) for any \(x\). Consequently, \(d_{\mathcal{D}_0}\) is a valid random policy.

Since the behavior policy is known, the proposed doubly-robust estimator corresponds to an unbiased estimator for \(\mathcal{L}^{-1} \sum_{\ell=1}^\mathcal{L} V(d_{\mathcal{D}_0})\). Using similar arguments in the causal inference literature on deriving the asymptotic property of doubly-robust estimators (Chernozhukov et al. 2017), we can show that
\[
\bar{V}(d) - \frac{1}{\mathcal{L}} \sum_{\ell=1}^\mathcal{L} V(d_{\mathcal{D}_0}) = O_p(n^{-1/2}).
\]
It suffices to show \( L^{-1} \sum_{\ell=1}^{L} \{ V(d_{\mathcal{G}[\ell]}) - V(d) \} = O_p \{ n^{-2\beta/(2\beta+p)} \log^8 n \} \), or equivalently, \( V(d_{\mathcal{G}[\ell]}) - V(d) = O_p \{ n^{-2\beta/(2\beta+p)} \log^8 n \} \).

Based on the results obtained in the first two parts, it follows from Cauchy-Schwarz inequality that

\[
\sum_{\mathcal{J} \in \mathcal{G}(\ell)} E\left[ \mathbb{I}(A \in \mathcal{J}) | Q(X, A) - q_{\mathcal{J}}^{(\ell)}(X) |^2 | X \right] \leq 2 \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E\left[ \mathbb{I}(A \in \mathcal{J}) | Q(X, A) - q_{\mathcal{J},0}(X) |^2 \right] + 2 \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E\left[ \mathbb{I}(A \in \mathcal{J}) | q_{\mathcal{J}}^{(\ell)}(X) - q_{\mathcal{J},0}(X) |^2 | X \right] \propto n^{-2\beta/(p+2\beta)} \log^8 n.
\]

(B.19)

Note that

\[
V(d_{\mathcal{G}[\ell]}) = \mathbb{E} \left[ \mathbb{I}(A \in \mathcal{J}) | Q(X, A) - q_{\mathcal{J}}^{(\ell)}(X) |^2 | X \right] \leq 2 \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E\left[ \mathbb{I}(A \in \mathcal{J}) | Q(X, A) - q_{\mathcal{J},0}(X) |^2 \right] + 2 \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E\left[ \mathbb{I}(A \in \mathcal{J}) | q_{\mathcal{J}}^{(\ell)}(X) - q_{\mathcal{J},0}(X) |^2 | X \right] \propto n^{-2\beta/(p+2\beta)} \log^8 n.
\]

It suffices to show

\[
V(d) = \sum_{\mathcal{J} \in \mathcal{O}_0} E q_{\mathcal{J}}(X) \mathbb{I}\{ d(X) \in \mathcal{J} \}.
\]

Similarly, we can show

\[
V(d) = \sum_{\mathcal{J} \in \mathcal{O}_0} E q_{\mathcal{J}}(X) \mathbb{I}\{ d(X) \in \mathcal{J} \}.
\]

It follows that

\[
| V(d_{\mathcal{G}[\ell]}) - V(d) | \leq \sum_{\mathcal{J} \in \mathcal{O}_0} E | q_{\mathcal{J}}(X) | \left| \mathbb{I}\{ d(X) \in \mathcal{J} \} \right| - \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E \mathbb{I}\{ d(X) \in \mathcal{J} \} \frac{b(\mathcal{J} \cap \mathcal{O}_0 | X)}{b(\mathcal{J} | X)}.
\]

As \( q_{\mathcal{J}} \) is uniformly bounded, the left-hand-side is upper bounded by

\[
\sum_{\mathcal{J} \in \mathcal{O}_0} E \left| \mathbb{I}\{ d(X) \in \mathcal{J} \} \right| - \sum_{\mathcal{J} \in \mathcal{G}(\ell)} E \mathbb{I}\{ d(X) \in \mathcal{J} \} \frac{b(\mathcal{J} \cap \mathcal{O}_0 | X)}{b(\mathcal{J} | X)}.
\]

(B.20)

Based on the results obtained in Part 2, for each \( \mathcal{J} \in \mathcal{O}_0 \), there exists some \( \mathcal{J}^{(\ell)} \) where the Lebesgue measure of the difference \( \mathcal{J} \cap (\mathcal{J}^{(\ell)})^c + \mathcal{J}^c \cap \mathcal{J}^{(\ell)} \) is upper bounded by \( O \{ n^{-2\beta/(2\beta+p)} \log^8 n \} \), with probability at least \( 1 - O(n^{-2}) \). The upper bound in (B.20) is \( O \{ n^{-2\beta/(2\beta+p)} \log^8 n \} \), under the positivity assumption and the assumption that \( \Pr(d(X) \in [\tau_0 - \epsilon, \tau_0 + \epsilon]) = O(\epsilon) \) for any \( \tau_0 \in J(\mathcal{O}_0) \) and sufficiently small \( \epsilon > 0 \). This completes the proof.
B.4.5 Proof of Theorem 3.4.2

We break the proof into two parts. In Part 1, we introduce an auxiliary lemma and present its proof. In Part 2, we derive the convergence rate of the proposed value estimator.

Part 1. We first introduce the following lemma.

**Lemma B.4.4.** For any interval $I \in \mathcal{I}(m)$ with $|I| \gg \gamma_n$ and any interval $I' \in \mathcal{D}(\ell)$ with $I \subseteq I'$, we have with probability approaching 1 that

$$E|q_{I,0}(X) - q_{I',0}(X)|^2 \leq \bar{C} |I|^{-1} \gamma_n,$$

for some constant $\bar{C} > 0$.

We next prove Lemma B.4.4. For a given interval $I' \in \mathcal{D}(\ell)$, the set of intervals $I$ considered in Lemma B.4.4 can be classified into the following three categories.

**Category 1:** $I = I'$. It is immediate to see that $q_I = q_{I'}$ and the assertion automatically holds.

**Category 2:** There exists another interval $I^* \in \mathcal{I}(m)$ that satisfies $I' = I^* \cup I$. Notice that the partition $\mathcal{D}(\ell)^* = \mathcal{D}(\ell) \cup \{I^*\} \cup \mathcal{I} - \{I'\}$ forms another partition. By definition, we have

$$\frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \sum_{\mathcal{I} \in \mathcal{D}(\ell)^*} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \tilde{q}_{I^*}(X_i))^2 + \gamma_n |\mathcal{D}(\ell)^*| \geq \frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \sum_{\mathcal{I} \in \mathcal{D}(\ell)} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \tilde{q}_{I^*}(X_i))^2 + \gamma_n |\mathcal{D}(\ell)|,$$

and hence

$$\frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \mathbb{I}(A_i \in \mathcal{I})(Y_i - \tilde{q}_{I^*}(X_i))^2 + \frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \mathbb{I}(A_i \in \mathcal{I}^*)(Y_i - \tilde{q}_{I^*}(X_i))^2 \geq \frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \mathbb{I}(A_i \in \mathcal{I}^*)(Y_i - \tilde{q}_{I^*}(X_i))^2 - \gamma_n.$$

It follows from the definition of $\tilde{q}_{I^*}$ that

$$\frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \mathbb{I}(A_i \in \mathcal{I}^*) (Y_i - \tilde{q}_{I^*}(X_i))^2 \leq \frac{1}{|L_c^\ell|} \sum_{i \in L_c^\ell} \mathbb{I}(A_i \in \mathcal{I}^*) (Y_i - \tilde{q}_{I^*}(X_i))^2.$$
Therefore, we obtain

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 \geq \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 - \gamma_n. \]  

(B.21)

**Category 3:** There exist two intervals \( \mathcal{I}^*, \mathcal{I}^{**} \in \mathcal{I}(m) \) that satisfy \( \mathcal{I}' = \mathcal{I}^* \cup \mathcal{I} \cup \mathcal{I}^{**} \). Using similar arguments in proving (B.21), we can show that

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 \geq \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 - 2\gamma_n. \]

Hence, regardless of whether \( \mathcal{I} \) belongs to Category 2, or it belongs to Category 3, we have

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 \geq \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 - 2\gamma_n. \]  

(B.22)

Notice that for any interval \( \mathcal{I}_0 \),

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I}_0)\{ Y_i - \hat{q}_{x_0}(X_i) \}^2 - E[\mathbb{I}(A \in \mathcal{I}_0)\{ Y - \hat{q}_{x_0}(Y) \}^2|\{ O_i \}_{i \in L'_\ell}] \]

\[ = \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I}_0)\{ \hat{q}_{x_0}(X_i) - q_{x_0,0}(X_i) \} \{ q_{x_0,0}(X_i) - \hat{q}_{x_0,0}(X_i) \}^2 \]

\[ + \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I}_0)\{ Y_i - \hat{q}_{x_0}(X_i) \}^2 - E[\mathbb{I}(A \in \mathcal{I}_0)\{ \hat{q}_{x_0}(X_i) - \hat{q}_{x_0,0}(X_i) \}^2|\{ O_i \}_{i \in L'_\ell}] \].

Using similar arguments in bounding the stochastic error term in Part 2 of the proof of Lemma B.4.1, we can show with probability approaching 1 that the right-hand-side is of the order \( O\{ n^{-2\beta/2\beta+p}\log^8 n \} \), for any \( \mathcal{I}_0 \in \mathcal{I}(m) \). As such, we obtain with probability approaching 1 that

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 = E[\mathbb{I}(A \in \mathcal{I})\{ Y - \hat{q}_{x'}(Y) \}^2|\{ O_i \}_{i \in L'_\ell}] \]

\[ + O(1)|\mathcal{I}|(n|\mathcal{I}|)^{-2\beta/2\beta+p}\log^8 n, \]

\[ \frac{1}{|L'_\ell|} \sum_{i \in L'_\ell} \mathbb{I}(A_i \in \mathcal{I})\{ Y_i - \hat{q}_{x'}(X_i) \}^2 = E[\mathbb{I}(A \in \mathcal{I})\{ Y - \hat{q}_{x'}(Y) \}^2|\{ O_i \}_{i \in L'_\ell}] \]

\[ + O(1)|\mathcal{I}|(n|\mathcal{I}|)^{-2\beta/2\beta+p}\log^8 n, \]

where \( O(1) \) denotes some universal positive constant. Combining these together with (B.22)
Consider the first term on the right-hand-side. Note that $O_i$ bounded by
By Lemma B.4.1, Lemma B.4.3 and the positivity assumption, the right-hand-side is upper
By Cauchy-Schwarz inequality, the last term on the right-hand-side can be lower bounded
It follows that
Consider the first term on the right-hand-side. Note that
By Cauchy-Schwarz inequality, the last term on the right-hand-side can be lower bounded by
It follows that
and hence
By Lemma B.4.1, Lemma B.4.3 and the positivity assumption, the right-hand-side is upper bounded by $O(1)|\mathcal{J}||n|\mathcal{J}|^{-2\beta/(2\beta+p)}\log^8 n$ for some universal positive constant $O(1)$, with
Theorem 3.4.1, the bias is given by
\[ L \rightarrow \gamma R q \text{ can be represented as} \]
that
\[ I \]
uniformly for any \( I \) and \( I' \), or equivalently,
\[ E \left[ \frac{b(I|X)}{|I|} \{ q_x(X) - q_{x'}(X) \}^2 \{ O_i \}_{i \in I} \right] = \frac{4\gamma R}{|I|} + O(1) \{|n||I|\}^{-2\beta/(2\beta + p)} \log^8 n. \]
By the positivity assumption, we have with probability approaching 1 that
\[ E[\{|q_x(X) - q_{x'}(X)\}^2 \{ O_i \}_{i \in I}] = O(\gamma n|I|^{-1}) + O(\{|n||I|\}^{-2\beta/(2\beta + p)} \log^8 n), \]
uniformly for any \( I \) and \( I' \). The proof is hence completed by noting that \( \gamma n \) is at least of the order \( O(n^{-2\beta}/(2\beta + p)) \log^8 n \).

Part 2. Consider the bias of the proposed estimator first. Similar to Part 3 of the proof of Theorem 3.4.1, the bias is given by \( S^{-1} \sum_{t=1}^{\overline{S}} V(d_{\overline{Q}(0)} - V(d)). \) By definition,
\[ V(d_{\overline{Q}(0)}) - V(d) = \sum_{\mathcal{I} \in \overline{Q}(0)} \int_{\mathcal{I}} E\{Q(X, a)I(d(X) \in \mathcal{I}) \frac{p(a|x)}{b(I|X)} \} d a - E\{Q(X, d(X))\} \]
\[ = \sum_{\mathcal{I} \in \overline{Q}(0)} \int_{\mathcal{I}} E\{Q(X, a) - Q(X, d(X))\} \frac{p(a|x)}{b(I|X)} \} d a \]
\[ = \sum_{\mathcal{I} \in \overline{Q}(0)} E\{q_{x,0}(X) - Q\{X, d(X)\}\} I(d(X) \in \mathcal{I}). \]
It follows that
\[ |V(d_{\overline{Q}(0)}) - V(d)| \leq \sup_{\mathcal{I} \in \overline{Q}(0), a \in \mathcal{I}'} E|Q(X, a) - q_{x'}(X)|. \]

For any \( \mathcal{I}' \in \overline{Q}(0) \). Consider two separate cases, corresponding to \( \mathcal{I}' \leq \gamma n^{1/3} \) and \( \mathcal{I}' > \gamma n^{1/3} \), respectively.

In Case 1, it follows from the Lipschitz property of the conditional mean function \( Q \) that \( |Q(x, a_1) - Q(x, a_2)| \leq L \gamma n^{1/3} \), for any \( a_1, a_2 \in \mathcal{I}' \) and \( x \). By definition, the function \( q_{x'} \) can be represented as \( q_{x'}(x) = \int_{\mathcal{I}} Q(x, a) \omega(x, a) d a \) for some weight function \( \omega \) such that \( \int_{\mathcal{I}} \omega(x, a) d a = 1 \). It follows that the right-hand-side of (B.23) is upper bounded by \( L \gamma n^{1/3} \).

In Case 2, for any \( a \in \mathcal{I}' \), we can find an interval \( \mathcal{I} \subseteq \mathcal{I}' \), \( a \in \mathcal{I} \) with length proportional to \( \gamma n^{1/3} \). Using similar arguments in Case 1, we can show that \( |Q(x, a) - q_{x,0}(x)| \leq L \gamma n^{1/3} \). By
Lemma B.4.4 and the Cauchy-Schwarz inequality, we have

$$E|q_{\delta,0}(X) - q_{\delta',0}(X)| \leq \sqrt{C \gamma_n^{2/3}} = \tilde{C}^{1/2} \gamma_n^{1/3},$$

with probability approaching 1. It follows that the right-hand-side of (B.23) is upper bounded by $(L + \sqrt{C})\gamma_n^{1/3}$, with probability approaching 1.

As such, the bias of the proposed estimator is upper bounded by $(L + \sqrt{C})\gamma_n^{1/3}$, with probability approaching 1.

We next consider the standard deviation of our estimator. The proposed estimator is can be represented by $L^{-1} \sum_{\ell=1}^{\ell} \tilde{V}^{\ell}(d)$ where $\tilde{V}^{\ell}(d)$ is the value estimator constructed based on the samples in $\{O_i\}_{i \in \mathcal{L}_\ell}$. Since the propensity score function is known to us, each $\tilde{V}^{\ell}(d)$ is unbiased to $V(d_{\mathcal{G}(\ell)})$. Under the positivity assumption and the boundedness assumption on $Y$ and $\hat{q}_{\delta}$, the variance of $\tilde{V}^{\ell}(d)$ is upper bounded by $|\mathcal{L}_\ell|^{-1} \inf_{\delta \in \mathcal{G}(\ell)} |\mathcal{G}|^{-1}$. By Lemma B.4.3, it is upper bounded by $O(n^{-1}\gamma_n^{-1})$. As such, the standard deviation of our estimator is upper bounded by $O(n^{-1}\gamma_n^{-1})$.

As such, the convergence rate is given by $O_p(\gamma_n^{1/3} + n^{-1/2}\gamma_n^{-1/2})$. By setting $\gamma_n = n^{-3/5}$, the rate is given by $O_p(n^{-1/5})$. The proof is hence completed.
C.1 Estimation on Variances

In this section, we present the estimators for $\sigma_Y^2$, $\rho$, $\rho_R$, $\Sigma_M$, and $\Sigma_R$. Recall $\hat{\pi}_p$, $\hat{\pi}_U$, $\hat{\pi}$, $\hat{\mu}_p$, $\hat{\theta}$, and $\hat{r}$ are estimators for the propensity score functions $\pi_p$, $\pi_U$, and $\pi$, the conditional mean functions $\mu_p$ and $\theta$, and the posterior sampling probability $r$, respectively, using any parametric or nonparametric models such as Random Forest or Deep Learning. Our theoretical results still hold with these nonparametric estimators as long as the estimators have desired convergence rates (see results established in Wager and Athey (2018); Farrell et al. (2021)). To introduce the variance estimators, we first define the value functions at the individual level. Given a decision rule $d(\cdot)$, let the value for the $i$-th individual in terms of the outcome of interest as

$$
\hat{v}_p^{(i)}(d) := \frac{\mathbb{I}[A_{p,i} = d(X_{p,i})][Y_{p,i} - \hat{\mu}_p\{X_{p,i}, d(X_{p,i})\}]}{A_{p,i}\hat{\pi}_p(X_{p,i}) + (1 - A_{p,i})(1 - \hat{\pi}_p(X_{p,i}))} + \hat{\mu}_p\{X_{p,i}, d(X_{p,i})\},
$$

in the primary sample, for $i \in \{1, \cdots, N_p\}$. Similarly, the value for the $i$-th individual in terms
of intermediate outcomes are
\[
\hat{w}_p^{(i)}(d) := \frac{\mathbb{I}\{A_{P,i} = d(X_{P,i})\}[M_{P,i} - \hat{\theta}\{X_{P,i}, d(X_{P,i})\}]}{A_{P,i}\hat{\pi}_p(X_{P,i}) + (1-A_{P,i})[1-\hat{\pi}_p(X_{P,i})]} + \hat{\theta}\{X_{P,i}, d(X_{P,i})\},
\]
in the primary sample for \( i \in \{1, \cdots, N_p\} \), and
\[
\hat{w}_U^{(i)}(d) := \frac{\mathbb{I}\{A_{U,i} = d(X_{U,i})\}[M_{U,i} - \hat{\theta}\{X_{U,i}, d(X_{U,i})\}]}{A_{U,i}\hat{\pi}_U(X_{U,i}) + (1-A_{U,i})[1-\hat{\pi}_U(X_{U,i})]} + \hat{\theta}\{X_{U,i}, d(X_{U,i})\},
\]
in the auxiliary sample for \( i \in \{1, \cdots, N_U\} \), where \( \hat{w}_p^{(i)}(d) \) and \( \hat{w}_U^{(i)}(d) \) are \( s \times 1 \) vectors.

Following the results in (4.1) and Lemma 4.3.2, we propose to estimate \( \sigma^2_Y(\cdot), \rho(\cdot) \) and \( \Sigma_M(\cdot) \) by
\[
\tilde{\sigma}^2_Y(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{w}_p^{(i)}(d) - \hat{V}_p(d) \right\}^2,
\]
\( (C.1) \)
\[
\bar{\rho}(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{w}_p^{(i)}(d) - \hat{V}_p(d) \right\} \left\{ \hat{w}_U^{(i)}(d) - \hat{W}_U(d) \right\},
\]
\[
\Sigma_M(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{w}_p^{(i)}(d) - \hat{W}_p(d) \right\} \otimes \frac{1}{N_U} \sum_{i=1}^{N_U} \left\{ \hat{w}_U^{(i)}(d) - \hat{W}_U(d) \right\},
\]
where \( \eta \otimes \eta^\top \) for \( \eta \) as a vector.

Similarly, based on the results in Lemma 4.3.4, we define the rebalanced value for the \( i \)-th individual in terms of intermediate outcomes as
\[
\tilde{w}_1^{(i)}(d) := \frac{R_{i}}{\hat{r}\{X_i, d(X_i), M_i\}} \frac{\mathbb{I}\{A_i = d(X_i)\}[M_i - \hat{\theta}\{X_i, d(X_i)\}]}{A_i\hat{\pi}(X_i) + (1-A_i)[1-\hat{\pi}(X_i)]} + \hat{\theta}\{X_i, d(X_i)\},
\]
\[
\tilde{w}_0^{(i)}(d) := \frac{(1-R_i)}{1-\hat{r}\{X_i, d(X_i), M_i\}} \frac{\mathbb{I}\{A_i = d(X_i)\}[M_i - \hat{\theta}\{X_i, d(X_i)\}]}{A_i\hat{\pi}(X_i) + (1-A_i)[1-\hat{\pi}(X_i)]} + \hat{\theta}\{X_i, d(X_i)\},
\]
where \( \tilde{w}_0^{(i)}(d) \) and \( \tilde{w}_1^{(i)}(d) \) are \( s \times 1 \) vectors, for \( i = 1, \cdots, N_p+N_U \). Also, we define the correlated part for the \( i \)-th individual in terms of intermediate outcomes as
\[
\tilde{\psi}^{(i)}(d) := \frac{1}{\hat{r}\{X_i, d(X_i), M_i\}} \frac{\mathbb{I}\{A_i = d(X_i)\}}{A_i\hat{\pi}(X_i) + (1-A_i)[1-\hat{\pi}(X_i)]} [M_i - \hat{\theta}\{X_i, d(X_i)\}],
\]
where \( \tilde{\psi}^{(i)}(d) \) is \( s \times 1 \) vectors, for \( i = 1, \cdots, N_p \). We then propose to estimate \( \rho(\cdot) \) and \( \Sigma(\cdot) \)
by
\[
\hat{\rho}_R(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{v}_p^{(i)}(d) - \hat{V}_p(d) \right\} \sqrt{\frac{N_p}{n} \hat{\Sigma}^{-1}_R(d)},
\]
(C.2)
\[
\hat{\Sigma}_R(d) = \frac{1}{n} \sum_{i=1}^{n} \left[ (\hat{w}_1^{(i)}(d) - \hat{w}_0^{(i)}(d))^2 \right].
\]

### C.2 Extension of Iterative Policy Tree Search Algorithm

The iterative policy search algorithm also works for finding ODR under CODA-HE, by replacing the corresponding variance estimates from (C.1) with that from (C.2). Specifically, we can rewrite the calibrated value estimator in (4.5) as

\[
\hat{V}_R(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \hat{v}_p(d) - \sqrt{\frac{n}{N_p}} \hat{\rho}_R(d)^\top \hat{\Sigma}_R^{-1}(d) \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_1^{(i)}(d) - \hat{w}_0^{(i)}(d))
\]

This motivates the calibrated reward for the \( i \)-th individual in the heterogeneous case as

\[
\hat{v}_p^{(i)}(1) = \frac{n}{N_p} \hat{\rho}_R(d_p)^\top \hat{\Sigma}_R^{-1}(d_p) \left[ \frac{N_p}{n} (\hat{w}_1^{(i)}(1) - \hat{w}_0^{(i)}(1)) + \hat{\Delta}(1) \right],
\]

under treatment 1, where \( \hat{\Delta}(1) = n^{-1} \sum_{i=N_p+1}^{n} (\hat{w}_1^{(i)}(1) - \hat{w}_0^{(i)}(1)) \), and

\[
\hat{v}_p^{(i)}(0) = \frac{n}{N_p} \hat{\rho}_R(d_p)^\top \hat{\Sigma}_R^{-1}(d_p) \left[ \frac{N_p}{n} (\hat{w}_1^{(i)}(0) - \hat{w}_0^{(i)}(0)) + \hat{\Delta}(0) \right],
\]

under treatment 0, where \( \hat{\Delta}(0) = n^{-1} \sum_{i=N_p+1}^{n} (\hat{w}_1^{(i)}(0) - \hat{w}_0^{(i)}(0)) \). Then, using similar steps in the iterative policy tree search algorithm for CODA-HO, we can find ODR under CODA-HE.

We next extent the iterative policy tree search to parametric decision rules with an illustration in homogenous case. Suppose the decision rule \( d(\cdot) \) relies on a model parameter \( \beta \), denoted as \( d(\cdot) \equiv d(\cdot; \beta) \). We use a shorthand to write \( V(d) \) as \( V(\beta) \), and define \( \beta_0 = \arg \max_{\beta} V(\beta) \). Thus, the value for the primary outcome of interest under the true ODR \( d(\cdot; \beta_0) \) is defined as \( V(\beta_0) \). Suppose the decision rule takes a form as \( d(X; \beta) \equiv \mathbb{I}[g(X) \top \beta > 0] \), where \( g(\cdot) \) is an unknown function and \( \mathbb{I}(\cdot) \) is the indicator function. We use \( \phi_X(\cdot) \) to denote a set of basis functions of baseline covariates with length \( q \), which are rich enough
to approximate the underlying function $g(\cdot)$. Thus, the decision rule is found within a class of $I\{\phi_X(X)^\top \beta > 0\}$, denoted as the class $\Pi_2$. Here, for notational simplicity, we include 1 in $\phi_X(\cdot)$ so that the parameter $\beta \in \mathbb{R}^{q+1}$.

Suppose the decision rule $d(\cdot)$ falls in class $\Pi_2$ that relies on a parametric model with parameters $\beta$. We use shorthands to write $\rho(d)$ as $\rho(\beta)$, $\Sigma_M(d)$ as $\Sigma_M(\beta)$, $\tilde{V}_p(d)$ as $\tilde{V}_p(\beta)$, $\tilde{W}_p(d)$ as $\tilde{W}_p(\beta)$, and $\tilde{W}_U(d)$ as $\tilde{W}_U(\beta)$, respectively. Then, the calibrated value estimator for $V(\beta)$ can be constructed by

$$\tilde{V}(\beta) = \tilde{V}_p(\beta) - \tilde{\rho}(\beta)^\top \Sigma_M^{-1}(\beta)\{\tilde{W}_p(\beta) - \tilde{W}_U(\beta)\},$$

where $\tilde{\rho}(\beta)$ is the estimator for $\rho(\beta)$, and $\Sigma_M(\beta)$ is the estimator for $\Sigma_M(\beta)$. We apply a similar iterative-updating procedure discussed above for CODA-HO but within parametric decision rules.

**Step 1**. Find the ODR in the primary sample by $\hat{\beta}_p = \arg\max_{\beta} \tilde{V}_p(\beta)$ as an initial decision rule. This can be solved using any global optimization algorithm, such as the heuristic algorithm provided in the R package rgenoud.

**Step 2**. Estimate covariances $\rho(\beta)$ and $\Sigma_M(\beta)$ based on (C.1) with $\beta = \hat{\beta}_p$. Then, we search the ODR based on the calibrated value estimator within the class $\Pi_2$ that maximizes

$$\tilde{V}(\beta) = \tilde{V}_p(\beta) - \tilde{\rho}(\beta)^\top \Sigma_M^{-1}(\beta)\{\tilde{W}_p(\beta) - \tilde{W}_U(\beta)\}.$$  

**Step 3**. Repeat step 2 for $k = 1, \cdots, K$, by replacing the previous estimated $\hat{\beta}^{(k-1)} (\hat{\beta}^{(0)} = \hat{\beta}_p)$ with the new estimated $\hat{\beta}^{(k)}$ until the number of iterations achieves $K$ or $\|\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}\|_2 < \delta$ where $\delta$ is a pre-specified threshold and $\|\cdot\|_2$ is the $L_2$ norm.

### C.3 Investigation with Multiple Intermediate Outcomes

We next consider the dimension of covariates as $r = 10$ and the dimension of intermediate outcomes as $s = 2$. The datasets are generated from the following three scenarios, respectively.

**Scenario 3**:

\[
\begin{cases}
U^M(X) = \begin{bmatrix} X^{(1)} + 2X^{(2)} \\ 0 \end{bmatrix}, & C^M(X) = \begin{bmatrix} X^{(1)} \times X^{(2)} \\ 0 \end{bmatrix} \\
U^Y(X) = 2X^{(1)} + X^{(2)}, & C^Y(X) = 2X^{(1)} \times X^{(2)}
\end{cases}
\]  

217
Table C.1: Empirical results of the proposed CODA-HO method in comparison to the ODR based on the primary sample solely under Scenarios 3 to 5.

<table>
<thead>
<tr>
<th>Scen.</th>
<th>Method (Rule)</th>
<th>CODA ($d^{opt}$)</th>
<th>CODA ($\hat{d}$)</th>
<th>ODR ($d^{opt}$)</th>
<th>ODR ($\hat{d}_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_p =$</td>
<td>500</td>
<td>1000</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>S3</td>
<td>Estimated $V(\cdot)$</td>
<td>0.983 0.986</td>
<td>1.037 1.021</td>
<td>0.984 0.980</td>
<td>1.072 1.038</td>
</tr>
<tr>
<td></td>
<td>$SD[V(\cdot)]$</td>
<td>0.130 0.093</td>
<td>0.128 0.093</td>
<td>0.182 0.126</td>
<td>0.180 0.125</td>
</tr>
<tr>
<td></td>
<td>$E[\sigma]$</td>
<td>0.130 0.096</td>
<td>0.129 0.095</td>
<td>0.182 0.128</td>
<td>0.181 0.128</td>
</tr>
<tr>
<td></td>
<td>Coverage Probabilities</td>
<td>94.6% 95.8%</td>
<td>95.2% 94.8%</td>
<td>94.6% 94.6%</td>
<td>92.2% 94.8%</td>
</tr>
<tr>
<td></td>
<td>Improved Efficiency</td>
<td>28.6% 25.0%</td>
<td>28.7% 25.8%</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>S4</td>
<td>Estimated $V(\cdot)$</td>
<td>0.980 0.984</td>
<td>1.037 1.022</td>
<td>0.984 0.980</td>
<td>1.072 1.038</td>
</tr>
<tr>
<td></td>
<td>$SD[V(\cdot)]$</td>
<td>0.148 0.104</td>
<td>0.145 0.103</td>
<td>0.182 0.126</td>
<td>0.180 0.125</td>
</tr>
<tr>
<td></td>
<td>$E[\sigma]$</td>
<td>0.148 0.107</td>
<td>0.148 0.107</td>
<td>0.182 0.128</td>
<td>0.181 0.128</td>
</tr>
<tr>
<td></td>
<td>Coverage Probabilities</td>
<td>95.0% 96.6%</td>
<td>95.8% 95.4%</td>
<td>94.6% 94.6%</td>
<td>92.2% 94.8%</td>
</tr>
<tr>
<td></td>
<td>Improved Efficiency</td>
<td>18.7% 16.4%</td>
<td>18.2% 16.4%</td>
<td>/ /</td>
<td>/ /</td>
</tr>
<tr>
<td>S5</td>
<td>Estimated $V(\cdot)$</td>
<td>1.898 1.895</td>
<td>1.977 1.948</td>
<td>1.898 1.889</td>
<td>2.004 1.962</td>
</tr>
<tr>
<td></td>
<td>$SD[V(\cdot)]$</td>
<td>0.116 0.083</td>
<td>0.113 0.080</td>
<td>0.147 0.102</td>
<td>0.142 0.099</td>
</tr>
<tr>
<td></td>
<td>$E[\sigma]$</td>
<td>0.116 0.084</td>
<td>0.116 0.085</td>
<td>0.150 0.106</td>
<td>0.150 0.106</td>
</tr>
<tr>
<td></td>
<td>Coverage Probabilities</td>
<td>94.8% 96.0%</td>
<td>92.2% 93.8%</td>
<td>95.2% 95.6%</td>
<td>91.0% 92.2%</td>
</tr>
<tr>
<td></td>
<td>Improved Efficiency</td>
<td>22.7% 20.8%</td>
<td>22.7% 19.8%</td>
<td>/ /</td>
<td>/ /</td>
</tr>
</tbody>
</table>

Scenario 4:
\[
\begin{align*}
U^M(X) &= \begin{bmatrix} 0.5\{X^{(1)}\}^2 + 2X^{(2)}, & X^M(X) = \begin{bmatrix} \{X^{(1)}\} \times X^{(2)}; \\
0, & 0, \end{bmatrix} \\
U^Y(X) &= 2X^{(1)} + X^{(2)}, C^Y(X) = 2X^{(1)} \times X^{(2)}. \end{align*}
\]

Scenario 5:
\[
\begin{align*}
U^M(X) &= \begin{bmatrix} X^{(1)} + 2X^{(2)}, & X^M(X) = \begin{bmatrix} X^{(1)} \times X^{(2)}, \\
0.5\{X^{(1)}\}^2 + 2X^{(2)}, & X^{(1)} \times X^{(2)} \end{bmatrix}; \\
U^Y(X) &= 2\cos\{X^{(1)}\} + X^{(2)}, C^Y(X) = 2X^{(1)} \times X^{(2)} \end{align*}
\]

The true ODR for Scenarios 3 to 5 is the same as $d^{opt}(X) = \mathbb{I}\{X^{(1)}X^{(2)} > 0\}$, with the true value $V(d^{opt})$ as 0.999 for Scenarios 3 and 4 while 1.909 for Scenario 5, based on Monte Carlo approximations. Using a similar procedure introduced in Section 4.5.1, we apply the proposed CODA-HO method, in comparison to the ODR method based on the primary sample solely. The empirical results are summarized in Table C.1 for Scenarios 3 to 5, aggregated over 500 replications.

It can be seen from Table C.1 that the proposed CODA-HO procedure performs reason-
ably better than the baseline procedure in terms of smaller variance under all scenarios. Specifically, in Scenario 3 with the baseline function linear in $X$, CODA-HO achieves a standard deviation of 0.095, against the larger standard deviation of 0.128 under the ODR based on the primary sample solely, with improved efficiency as 25.8%, under $N_p = 1000$. In Scenarios 4 and 5 with more complex non-linear baseline functions, CODA-HO outperforms the original ODR method by reducing the standard deviation as 16.4% and 19.8%, respectively, under $N_p = 1000$. In addition, the estimated value function under the estimated ODR obtained by CODA-HO achieves better coverage probabilities in comparison to the corresponding estimators obtained using the primary sample solely under all scenarios when the sample size is small, $N_p = 500$, indicating a stronger capacity of the proposed method in handling high-dimensional covariates by incorporating more samples with multiple mediators.

C.4 Technical Proofs

We give technical proofs for the established lemmas and theorems in this section.

C.4.1 Proof of Lemma 4.3.1

The proof of Lemma 4.3.1 consists of five steps as follows. We remark that the key ingredient of the proof lies in the law of iterated expectation together with assumptions (A1) and (A4) under the homogeneous baseline covariates, $X_p \sim X_U$.

(s1.) First, for $E\{M_p^*(d)\} = E[M_p^*(0)\{1 - d(X_p)\} + M_p^*(1)d(X_p)]$, taking its iterated expectation on $\{X_p, A_p\}$, we have

\[
E\{M_p^*(d)\} = E[M_p^*(0)\{1 - d(X_p)\} + M_p^*(1)d(X_p)] \\
= E\left[ E\left\{ M_p^*(0)\{1 - d(X_p)\} + M_p^*(1)d(X_p) \right\} \bigg| X_p = x, A_p = a \right] \\
= E\left[ E\left\{ M_p^*(0)\{1 - d(x)\} + M_p^*(1)d(x) \right\} \bigg| X_p = x, A_p = a \right].
\]

(s2.) By the assumption (A1) that $M_p = A_pM_p^*(1) + (1 - A_p)M_p^*(0)$, with the fact $A_pA_p = A_p$ and $A_p(1 - A_p) = 0$ for $A_p \in \{0, 1\}$, we have

\[
A_pM_p = A_p\{A_pM_p^*(1) + (1 - A_p)M_p^*(0)\} = M_p^*(1).
\]
As such, we represent $M_P^*(0)$ as

\[(1 - A_P)M_P = (1 - A_P)\{A_PM_P^*(1) + (1 - A_P)M_P^*(0)\} = M_P^*(0).\]

(s3.) By replacing $M_P^*(1)$ and $M_P^*(0)$ in the value function with results in (s2.), we have

\[E[M_p^*(d)] = E\left\{(1 - A_P)M_P\{1 - d(x)\} + A_PM_Pd(x) \left| X_p = x, A_P = a \right.\right\} = E\left\{(1 - A_P)M_P\{1 - d(x)\} + A_PM_Pd(x) \left| X_p = x, A_P = a \right.\right\} = E\left\{(1 - A_P)(1 - d(x)) + ad(x) \left| M_P \left| X_p = x, A_P = a \right.\right.\right\}.

(s4.) By the assumption (A4) that $E[M_p|X_p = x, A_p = a] = E[M_u|X_u = x, A_u = a]$, we have

\[E[M_p^*(d)] = E\left\{(1 - A_P)(1 - d(x)) + ad(x) \left| M_p \left| X_p = x, A_p = a \right.\right.\right\} = E\left\{(1 - A_P)(1 - d(x)) + ad(x) \left| M_u \left| X_u = x, A_u = a \right.\right.\right\},\]

where the last step is valid since the expectation is taking over the same baseline distributions ($X_p \sim X_u$) on two sides.

(s5.) Finally, the inverse procedure of steps (s3.) to (s1.) on the results in the step (s4.) leads to

\[E[M_p^*(d)] = E[M_p^*(0)\{1 - d(X_u)\} + M_u^*(1)d(X_u)] = E[M_u^*(d)].\]

The proof is hence completed.

### C.4.2 Proof of Lemma 4.3.2

We detail the proof for Lemma 4.3.2 in this section. First, the term $\sqrt{N_p[\hat{W}_p(d) - \hat{W}_u(d)]}$ can be decomposed by

\[\sqrt{N_p[\hat{W}_p(d) - \hat{W}_u(d)]} = \sqrt{N_p[\hat{W}_p(d) - W(d) + W(d) - \hat{W}_u(d)]} = \sqrt{N_p[\hat{W}_p(d) - W(d)]} - \sqrt{N_p[\hat{W}_u(d) - W(d)]},\]

where $\sqrt{t} = \sqrt{N_p/N_u}$ denotes the square root of the sample ratio.
We next show the third line in (C.3) is asymptotically normal with mean zero. To do this, we define the independent and identically distributed double robust terms of intermediate outcomes in two samples as

\[
\mathbf{w}_p^{(i)}(d) := \frac{\mathbbm{1}\{d(X_{P,i})\}[M_{P,i} - \theta \{X_{P,i}, d(X_{P,i})\}]}{A_{P,i} \pi_p(X_{P,i}) + (1 - A_{P,i})\{1 - \pi_p(X_{P,i})\}} + \theta \{X_{P,i}, d(X_{P,i})\},
\]

and

\[
\mathbf{w}_U^{(i)}(d) := \frac{\mathbbm{1}\{d(X_{U,i})\}[M_{U,i} - \theta \{X_{U,i}, d(X_{U,i})\}]}{A_{U,i} \pi_U(X_{U,i}) + (1 - A_{U,i})\{1 - \pi_U(X_{U,i})\}} + \theta \{X_{U,i}, d(X_{U,i})\},
\]

where \(\pi_p\) and \(\pi_U\) are the true propensity score function in two samples, and \(\theta\) is the true conditional mean function of the intermediate outcome given the covariates and the treatment.

Following the proof of Theorem 1 in the appendix B of Rai (2018), under assumptions (A6) and (A7. i, ii, and iii), given a decision rule \(d(\cdot)\) that satisfies the assumption (A5), we have

\[
\sqrt{N_p} \left[ \mathbf{\bar{W}}_p(d) - \mathbf{\bar{W}}_p(d) \right] = \sqrt{N_p} \left[ \mathbf{\bar{W}}_p(d) - W(d) \right] - \sqrt{t} \sqrt{N_U} \left[ \mathbf{\bar{W}}_U(d) - W(d) \right],
\]

(C.5)

Since \(t \in (0, +\infty)\).

Based on the assumption (A4), \(X_p \sim X_U\), and the central limit theorem, we have

\[
\sqrt{N_p} \left[ \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{w}_p^{(i)}(d) - W(d) \right] \to N_s \left\{ 0_s, \Sigma_p(d) \right\},
\]

(C.6)

\[
\text{and} \quad \sqrt{N_U} \left[ \frac{1}{N_U} \sum_{i=1}^{N_U} \mathbf{w}_U^{(i)}(d) - W(d) \right] \to N_s \left\{ 0_s, \Sigma_U(d) \right\},
\]

(C.7)

where \(0_s\) is \(s\)-dimensional zero vector, \(\Sigma_p\) and \(\Sigma_U\) are \(s \times s\) matrices, and \(N_s(\cdot, \cdot)\) is the \(s\)-dimensional multivariate normal distribution.

Notice that the two samples \((P\) and \(U\)) are independently collected from two separate studies, and thus the independent and identically distributed double robust terms \(\{\mathbf{w}_p^{(i)}(d)\}_{1 \leq i \leq N_p}\) are independent of \(\{\mathbf{w}_U^{(i)}(d)\}_{1 \leq i \leq N_U}\). Hence, by noting the fact that the linear
combination of two independent random variables having a normal distribution also
has a normal distribution, based on (C.5), (C.6), and (C.7), by Slutsky’s theorem, with
\[ T = \lim_{N_P \to +\infty} t < +\infty, \]
we have
\[ p_{N_P} \xrightarrow{c} cW_P(d) - cW_U(d) \xrightarrow{c} N_s\{0, \Sigma_M(d)\}, \]  
where \( \Sigma_M(d) = \Sigma_P(d) + T\Sigma_U(d) \) is a \( s \times s \) matrix. The proof is hence completed.

### C.4.3 Proof of Lemma 4.3.3

We detail the proof for Lemma 4.3.3 in this section. We focus on proving the asymptotic
normality of \( p_{n}cW_1(d) \), and the results for \( p_{n}cW_0(d) \) can be shown in a similar manner. Let
\[
\tilde{Z}_i = \frac{R_i}{r\{X_i, d(X_i), M_i\} A_i \tilde{\pi}(X_i) + (1 - A_i)\{1 - \tilde{\pi}(X_i)\}} \left[ M_i - \tilde{\theta} \{X_i, d(X_i)\} \right] + \tilde{\theta} \{X_i, d(X_i)\},
\]
for \( i = 1, \cdots, n \), and thus \( \tilde{W}_i(d) = 1/n \sum_{i=1}^{n} \tilde{Z}_i \). Denote its counterpart as
\[
Z_i = \frac{R_i}{r\{X_i, d(X_i), M_i\} A_i \pi(X_i) + (1 - A_i)\{1 - \pi(X_i)\}} \left[ M_i - \theta \{X_i, d(X_i)\} \right] + \theta \{X_i, d(X_i)\}.
\]
It is immediate from the central limit theorem that
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \xrightarrow{c} N_s\{W^*(d), \Sigma_1(d)\}, \tag{C.9}
\]
where \( \Sigma_1 \) is a \( s \times s \) matrice presenting the asymptotic covariance matrice, and
\[
W^*(d) = E(Z_i) = \int E\{M|d(X), X\}\{P(R = 1)f(E, X) + P(R = 0)f(U, X)\}dX.
\]
According to (C.9), to show
\[
\sqrt{n}\left\{ \tilde{W}_i(d) - W^*(d) \right\} \xrightarrow{c} N_s\{0, \Sigma_1(d)\}, \tag{C.10}
\]
it is sufficient to show
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{Z}_i - Z_i) = o_p(1). \tag{C.11}
\]
We next bound the rest of this section is focusing on proving (C.11), by the following two steps.

To this end, we define a middle term to assist our derivation as

$$\hat{Z}_i = \frac{R_i}{r[X_i, d(X_i), M_i]} \frac{\mathbb{I}[A_i = d(X_i)]}{A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]} \left[ M_i - \tilde{\theta} \{X_i, d(X_i)\} \right] + \tilde{\theta} \{X_i, d(X_i)\}.$$

Combining (C.12) and (C.13) yields (C.11), and thus (C.10) is proved.

We focus on proving (C.12) first. Since

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{Z}_i - \tilde{Z}_i) = o_p(1),$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{Z}_i - Z_i) = o_p(1).$$

Combining (C.12) and (C.13) yields (C.11), and thus (C.10) is proved.

We focus on proving (C.12) first. Since

$$\hat{Z}_i - \tilde{Z}_i = \left\{ \frac{R_i}{\tilde{r} \{X_i, d(X_i), M_i\}} - \frac{R_i}{r \{X_i, d(X_i), M_i\}} \right\} \frac{\mathbb{I}[A_i = d(X_i)]}{A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]} \left[ M_i - \tilde{\theta} \{X_i, d(X_i)\} \right] + \tilde{\theta} \{X_i, d(X_i)\},$$

where $r_i \equiv r \{X_i, d(X_i), M_i\}$ and $\tilde{r}_i \equiv \tilde{r} \{X_i, d(X_i), M_i\}$. We can further decompose (C.14) by

$$\hat{Z}_i - \tilde{Z}_i = \left\{ \frac{R_i (r_i - \tilde{r}_i)}{\tilde{r}_i r_i} \right\} \frac{\mathbb{I}[A_i = d(X_i)]}{A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]} \left[ M_i - \theta \{X_i, d(X_i)\} \right] + \tilde{\theta} \{X_i, d(X_i)\}$$

$$+ \left[ \frac{r_i - \tilde{r}_i}{\tilde{r}_i r_i} \right] \frac{\mathbb{I}[A_i = d(X_i)]}{A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]} \left[ \frac{R_i A_i}{A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]} \right].$$

We next bound $\omega_{1,i}$ and $\omega_{2,i}$ using empirical process, respectively. Here, by definition, we have $\theta \{X_i, d(X_i)\} = E[M_i \mid X_i, d(X_i)]$, thus,

$$\sqrt{n}E_n \omega_{1,i} \leq \sqrt{n}E_n C_1 [M_i - \theta \{X_i, d(X_i)\}] = o_p(1),$$

where $E_n = \mathbb{E}[\mathbb{I}[A_i = d(X_i)] / A_i E[X_i] + (1 - A_i)[1 - \tilde{\pi}(X_i)]]$. 

223
where $C_1$ is the bound of
\[
R_i(r_i - \hat{r}_i)[A_i = d(X_i)] \frac{I\{A_i = d(X_i)\}[M_i - \theta \{X_i, d(X_i)\}]}{\hat{A}_i\hat{\pi}(X_i) + (1 - \hat{A}_i)[1 - \hat{\pi}(X_i)]} + \theta \{X_i, d(X_i)\}
\]

under the positivity assumption.

Also, by assumptions (A7. v), we have
\[
\sqrt{n}E_n\omega_{2,i} \leq \sqrt{n}E_nC_2(r_i - \hat{r}_i)[\theta \{X_i, d(X_i)\} - \hat{\theta} \{X_i, d(X_i)\}] = o_p(1), \tag{C.16}
\]
where $C_2$ is the bound of
\[
R_i[I\{A_i = d(X_i)\}] \frac{I\{A_i = d(X_i)\}[M_i - \theta \{X_i, d(X_i)\}]}{\hat{A}_i\hat{\pi}(X_i) + (1 - \hat{A}_i)[1 - \hat{\pi}(X_i)]}.
\]

Combining (C.15) with (C.16) yields (C.12).

Next, we show (C.13). Since
\[
\hat{Z}_i - Z_i = \frac{R_i}{r_i} \left\{ \frac{I\{A_i = d(X_i)\}[M_i - \theta \{X_i, d(X_i)\}]}{\hat{A}_i\hat{\pi}(X_i) + (1 - \hat{A}_i)[1 - \hat{\pi}(X_i)]} + \theta \{X_i, d(X_i)\} \right\}_{\omega_3,i}
\]

\[
- \frac{I\{A_i = d(X_i)\}[M_i - \hat{\theta} \{X_i, d(X_i)\}]}{\hat{A}_i\hat{\pi}(X_i) + (1 - \hat{A}_i)[1 - \hat{\pi}(X_i)]} - \hat{\theta} \{X_i, d(X_i)\} \right\}_{\omega_4,i}
\]

\[
+ \hat{\theta} \{X_i, d(X_i)\} - \theta \{X_i, d(X_i)\} \right\} + \theta \{X_i, d(X_i)\} - \hat{\theta} \{X_i, d(X_i)\}
\]

\[
= \frac{R_i}{r_i} \left\{ (\omega_{3,i} - \omega_{4,i}) + \hat{\theta} \{X_i, d(X_i)\} - \theta \{X_i, d(X_i)\} \right\} + \theta \{X_i, d(X_i)\} - \hat{\theta} \{X_i, d(X_i)\}
\]

\[
= \frac{R_i}{r_i} \left\{ (\omega_{3,i} - \omega_{4,i}) + \left( \frac{R_i}{r_i} - 1 \right) \hat{\theta} \{X_i, d(X_i)\} - \theta \{X_i, d(X_i)\} \right\}
\]

where $\omega_{3,i} - \omega_{4,i}$ is the regular doubly robust estimator (Zhang et al. 2012b) for $i$-th subject minus its counterpart. Using the similar arguments in proving (C.15), under assumption (A7. iv), we can show that
\[
\sqrt{n}E_n\frac{R_i}{r_i}(\omega_{3,i} - \omega_{4,i}) = o_p(1). \tag{C.18}
\]
On the other hand, since \( r_i = P(R_i = 1|X_i, d(X_i), M_i) \), we have

\[
\sqrt{n} E_n \omega_{5,i} \leq \sqrt{n} E_n C_3 \left( \frac{R_i}{r_i} - 1 \right) = o_p(1),
\]

(C.19)

where \( C_3 \) is the bound of \( \hat{\theta} \{X_i, d(X_i)\} - \theta \{X_i, d(X_i)\} \). This together with (C.18) yields (C.13). Thus, we complete the proof of (C.10).

### C.4.4 Proof of Lemma 4.3.4

The proof for Lemma 4.3.4 is a direct result from Lemma 4.3.3. Recall we have

\[
\sqrt{n} \left\{ \hat{\psi}_i(d) - W^*(d) \right\} \sim N_s \left\{ 0_s, \Sigma_1(d) \right\}, \text{ and } \sqrt{n} \left\{ \hat{\psi}_0(d) - W^*(d) \right\} \sim N_s \left\{ 0_s, \Sigma_0(d) \right\}.
\]

Based on the property of normal distribution, we have

\[
\sqrt{n} \left\{ \hat{\psi}_i(d) - \hat{\psi}_0(d) \right\} \sim N_s \left\{ 0_s, \Sigma_R(d) \right\},
\]

(C.20)

where \( \Sigma_R(d) \) is a \( s \times s \) asymptotic covariance matrix for term \( \psi^{(i)}_1(d) - \psi^{(i)}_0(d) \), where

\[
\psi^{(i)}_1(d) = \frac{R_i}{r\{X_i, d(X_i), M_i\}} \frac{I\{A_i = d(X_i)\}}{A_i \pi(X_i) + (1 - A_i)\{1 - \pi(X_i)\}} [M_i - \theta \{X_i, d(X_i)\}],
\]

and

\[
\psi^{(i)}_0(d) = \frac{(1 - R_i)}{1 - r\{X_i, d(X_i), M_i\}} \frac{I\{A_i = d(X_i)\}}{A_i \pi(X_i) + (1 - A_i)\{1 - \pi(X_i)\}} [M_i - \theta \{X_i, d(X_i)\}].
\]

Here, \( \psi^{(i)}_1(d) \) is non-zero if \( R = 1 \) (from the \( P \) sample), and \( \psi^{(i)}_0(d) \) is non-zero if \( R = 0 \) (from the \( U \) sample). Noticing that the two samples (\( P \) and \( U \)) are independently collected from two separate studies, we have

\[
\Sigma_R(d) = \Sigma_{1P}(d) + \Sigma_{0U}(d),
\]

where \( \Sigma_{1P}(d) = \text{Var}\{\psi^{(i)}_1(d)\} \) and \( \Sigma_{0U}(d) = \text{Var}\{\psi^{(i)}_0(d)\} \) are \( s \times s \) asymptotic covariance matrices. The proof is hence completed.
C.4.5 Proof of Theorem 4.4.1

The proof of Theorem 4.4.1 consists of seven parts. We show the consistency of each proposed estimator in each part.

Proof of results (i):

We first show the theoretical form of $\sigma^2_Y(d)$ and then prove the consistency of $\bar{\sigma}^2_Y(d)$ to $\sigma^2_Y(d)$. Define the independent and identically distributed double robust term of the primary outcome of interest in the primary sample as

$$v_p^{(i)}(d) := \frac{1}{N_p} \sum_{i=1}^{N_p} v_p^{(i)}(d) - V(d),$$

where $\pi_p$ and $\mu_p$ are the true propensity score function and the true conditional mean function of the primary outcome of interest in the primary sample, respectively.

Similarly, following the proof of Theorem 1 in the appendix B of Rai (2018), under assumptions (A6) and (A7), given a decision rule $d(\cdot)$ that satisfies the assumption (A5), we have

$$\sqrt{N_p} [\bar{V}_p(d) - V(d)] = \sqrt{N_p} [\frac{1}{N_p} \sum_{i=1}^{N_p} v_p^{(i)}(d) - V(d)] + o_p(1) \sim N\{0, \sigma^2_Y(d)\}, \quad (C.22)$$

where $\sigma^2_Y(d) = E[\{v_p(d) - V(d)\}^2]$ and

$$\sigma^2_Y(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \{v_p^{(i)}(d) - V(d)\}^2 \xrightarrow{p} \sigma^2_Y(d). \quad (C.23)$$

Thus, by the weak law of large number, it is immediate that

$$\frac{1}{N_p} \sum_{i=1}^{N_p} \{v_p^{(i)}(d) - V(d)\}^2 \xrightarrow{p} \sigma^2_Y(d). \quad (C.23)$$
We next decompose the proposed variance estimator by

\[
\hat{\sigma}^2_Y(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{v}_p(i) - \bar{V}_p(d) \right\}^2
\]

(C.24)

\[
= \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{v}_p(i) - v_p(i) + v_p(i) - V(d) + V(d) - \bar{V}_p(d) \right\}^2.
\]

According to (C.22), we have \( \bar{V}_p(d) = V(d) + o_p(1) \). Due to the assumption (A6) that covariates and outcomes are bounded, the above (C.24) yields

\[
\hat{\sigma}^2_Y(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{v}_p(i) - v_p(i) + v_p(i) - V(d) + o_p(1) \right\}^2
\]

\[
= \frac{1}{N_p} \sum_{i=1}^{N_p} \left[ \left\{ v_p(i) - V(d) \right\}^2 + \left\{ \hat{v}_p(i) - v_p(i) \right\} \left\{ \hat{v}_p(i) - v_p(i) - 2V(d) \right\} \right] + o_p(1).
\]

By noticing the term \( \left\{ \hat{v}_p(i) + v_p(i) - 2V(d) \right\} \) is bounded owing to the assumption (A6), based on the results in (C.21), we have \( N_p^{-1} \sum_{i=1}^{N_p} \eta_0 = o_p(1) \). This gives

\[
\hat{\sigma}^2_Y(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \hat{v}_p(i) - \bar{V}_p(d) \right\}^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ v_p(i) - V(d) \right\}^2 + o_p(1).
\]

(C.25)

Combining (C.25) with (C.23) yields \( \hat{\sigma}^2_Y(d) = \sigma^2_Y(d) + o_p(1) \). The proof of results (i) is hence completed.

**Proof of results (ii):**

Similarly, we first show the theoretical form of \( \rho(d) \) and then prove the consistency of \( \hat{\rho}(d) \) to \( \rho(d) \). Aware that the joint distribution of normal random variables is still normal, we have the following results based on Lemma 4.1 and Lemma 4.3.2.

\[
\sqrt{N_p} \left[ \begin{array}{c} \bar{V}_p(d) - V(d) \\ \bar{W}_p(d) - \bar{W}_{ij}(d) \end{array} \right] \overset{d}{\sim} N_{s+1} \left\{ 0_{s+1}, \left[ \begin{array}{c} \sigma^2_Y(d), \rho(d) \end{array} \right] \right\}, \quad \forall d(\cdot),
\]

where \( \rho(d) \) is the \( s \times 1 \) asymptotic covariance vector between the value estimator of the outcome of interest in the primary sample and the differences of the value estimators of intermediate outcomes between two samples.
Recall the results in (C.5) and (C.22) that
\[
\sqrt{N_p} \left[ \bar{V}_p(d) - V(d) \right] = \sqrt{N_p} \left[ - \frac{1}{N_p} \sum_{i=1}^{N_p} v_p^{(i)}(d) - V(d) \right] + o_p(1),
\] (C.26)
\[
\sqrt{N_p} \left[ \bar{W}_p(d) - \bar{W}_U(d) \right] = \sqrt{N_p} \left[ - \frac{1}{N_p} \sum_{i=1}^{N_p} w_p^{(i)}(d) - W(d) \right]
- \sqrt{t} \sqrt{N_U} \left[ - \frac{1}{N_U} \sum_{i=1}^{N_U} w_U^{(i)}(d) - W(d) \right] + o_p(1).
\]

The asymptotic covariance between \( \sqrt{N_p} \left[ \bar{V}_p(d) - V(d) \right] \) and \( \sqrt{N_p} \left[ \bar{W}_p(d) - \bar{W}_U(d) \right] \) comes from the correlation between \( \{v_p^{(i)}(d)\}_{1 \leq i \leq N_p} \) and \( \{w_p^{(i)}(d)\}_{1 \leq i \leq N_p}, \{w_U^{(i)}(d)\}_{1 \leq i \leq N_U} \).

By noticing that the two samples (\( P \) and \( U \)) are independently collected from two separate studies, the independent and identically distributed double robust terms \( \{v_p^{(i)}(d)\}_{1 \leq i \leq N_p} \) are independent of \( \{w_U^{(i)}(d)\}_{1 \leq i \leq N_U} \).

Using similar arguments in proving (C.8) and (C.22), we have
\[
\rho(d) = E \left[ \left\{ v_p(d) - V(d) \right\} \left\{ w_p(d) - W(d) \right\} \right],
\] (C.27)
where
\[
w_p(d) := \frac{\mathbb{I}\{d(X_p)\} \{M_p - \theta \{X_p, d(X_p)\}\}}{A_p \pi_p(X_p) + (1 - A_p) \{1 - \pi_p(X_p)\}} + \theta \{X_p, d(X_p)\}.
\]
Thus, by the weak law of large number, it is immediate that
\[
\frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ v_p^{(i)}(d) - V(d) \right\} \left\{ w_p^{(i)}(d) - W(d) \right\} \xrightarrow{p} \rho(d).
\] (C.28)

Next, to show the consistency of \( \hat{\rho}(d) \) to \( \rho(d) \), we decompose \( \hat{\rho}(d) \) as
\[
\hat{\rho}(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \bar{v}_p^{(i)}(d) - \bar{V}_p(d) \right\} \left\{ \bar{w}_p^{(i)}(d) - \bar{W}_p(d) \right\}
= \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \bar{v}_p^{(i)}(d) - v_p^{(i)}(d) + v_p^{(i)}(d) - V(d) + V(d) - \bar{V}_p(d) \right\}
\times \left\{ \bar{w}_p^{(i)}(d) - w_p^{(i)}(d) + w_p^{(i)}(d) - W(d) + W(d) - \bar{W}_p(d) \right\}.
\] (C.29)
Using similar arguments in proving (C.25), we can show

\[ \hat{\rho}(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \nu_p^{(i)}(d) - V(d) \right\} \left\{ \omega_p^{(i)}(d) - W(d) \right\} + o_p(1). \]

This together with (C.28) proves results (ii). The proof is hence completed.

**Proof of results (iii):**

We already established the theoretical form of \( \Sigma_M = \Sigma_p(d) + T\Sigma_U(d) \) in the proof of Lemma 4.3.2, where \( T = \lim_{N_p \to +\infty} t < +\infty \).

Following the similar arguments in establishing (C.23), we have

\[
\frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \omega_p^{(i)}(d) - W(d) \right\} \left\{ \omega_p^{(i)}(d) - W(d) \right\}^\top \to \Sigma_p(d), \tag{C.30}
\]

\[
\frac{1}{N_U} \sum_{i=1}^{N_U} \left\{ \omega_U^{(i)}(d) - W(d) \right\} \left\{ \omega_U^{(i)}(d) - W(d) \right\}^\top \to \Sigma_U(d).
\]

To show the consistency of \( \check{\Sigma}_M(d) \) to \( \Sigma_M \), we decompose \( \check{\Sigma}_M(d) \) as

\[
\check{\Sigma}_M(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \check{\omega}_p^{(i)}(d) - \omega_p^{(i)}(d) + \omega_p^{(i)}(d) - W(d) + W(d) - \check{W}_p(d) \right\}
\times \left\{ \check{\omega}_p^{(i)}(d) - \omega_p^{(i)}(d) + \omega_p^{(i)}(d) - W(d) + W(d) - \check{W}_p(d) \right\}^\top
\]

\[
+ t \frac{1}{N_U} \sum_{i=1}^{N_U} \left\{ \check{\omega}_U^{(i)}(d) - \omega_U^{(i)}(d) + \omega_U^{(i)}(d) - W(d) + W(d) + \check{W}_U(d) \right\}
\times \left\{ \check{\omega}_U^{(i)}(d) - \omega_U^{(i)}(d) + \omega_U^{(i)}(d) - W(d) + W(d) + \check{W}_U(d) \right\}^\top.
\tag{C.31}
\]

Using similar arguments in proving (C.25), we can show

\[
\check{\Sigma}_M(d) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ \omega_p^{(i)}(d) - W_p(d) \right\} \left\{ \omega_p^{(i)}(d) - W_p(d) \right\}^\top
\]

\[
+ t \frac{1}{N_U} \sum_{i=1}^{N_U} \left\{ \omega_U^{(i)}(d) - W_U(d) \right\} \left\{ \omega_U^{(i)}(d) - W_U(d) \right\}^\top + o_p(1).
\]

This together with (C.30) proves results (iii). The proof is hence completed.

**Proof of results (iv):**

The proof of this part can be shown using similar arguments as in the proof of results (ii). We first show the theoretical form of \( \rho_R(d) \) and then prove the consistency of \( \check{\rho}_R(d) \) to
we have $\rho_R(d)$ is the $s \times 1$ asymptotic correlation vector between the value estimator of the outcome of interest in the primary sample ($\sqrt{N_p}[\tilde{V}_p(d) - V(d)]$) and the new rebalanced value difference estimator of intermediate outcomes between two samples ($\sqrt{n}[\tilde{W}_i(d) - \tilde{W}_0(d)]$).

Recall the results in (C.5) and (C.20), we have

$$\sqrt{N_p}[\tilde{V}_p(d) - V(d)] = \sqrt{N_p}\left[ \frac{1}{N_p} \sum_{i=1}^{N_p} V^{(i)}_p(d) - V(d) \right] + o_p(1),$$

$$\sqrt{n}[\tilde{W}_i(d) - \tilde{W}_0(d)] = \sqrt{n}\left[ \frac{1}{n} \sum_{i=1}^{n} \left\{ \psi^{(i)}_1(d) - \psi^{(i)}_0(d) \right\} \right] + o_p(1),$$

where

$$\psi^{(i)}_1(d) = \frac{R_i}{r[X_i,d(X_i),M_i]} \frac{I[A_i = d(X_i)]}{A_i\pi(X_i) + (1 - A_i)(1 - \pi(X_i))} \left[ M_i - \theta(X_i,d(X_i)) \right],$$

and

$$\psi^{(i)}_0(d) = \frac{(1 - R_i)}{1 - r[X_i,d(X_i),M_i]} \frac{I[A_i = d(X_i)]}{A_i\pi(X_i) + (1 - A_i)(1 - \pi(X_i))} \left[ M_i - \theta(X_i,d(X_i)) \right].$$

The asymptotic covariance $\rho_R(d)$ thus comes from the correlation between $\{V^{(i)}_p(d)\}_{1 \leq i \leq N_p}$ and $\{\psi^{(i)}_1(d)\}_{1 \leq i \leq n}$, $\{\psi^{(i)}_0(d)\}_{1 \leq i \leq n}$. Here, $\psi^{(i)}_1(d)$ is non-zero if $R = 1$ (from the $P$ sample), and $\psi^{(i)}_0(d)$ is non-zero if $R = 0$ (from the $U$ sample). By noticing that the two samples ($P$ and $U$) are independently collected from two separate studies, the independent and identically distributed double robust terms $\{V^{(i)}_p(d)\}_{1 \leq i \leq N_p}$ are independent of $\{\psi^{(i)}_1(d)\}_{(N_p+1) \leq i \leq n}$, $\{\psi^{(i)}_0(d)\}_{1 \leq i \leq n}$.

Using similar arguments in proving (C.28), by the weak law of large number, it is immediate that

$$\frac{1}{N_p} \sum_{i=1}^{N_p} \left\{ V^{(i)}_p(d) - V(d) \right\} \frac{\sqrt{n}}{\sqrt{n}\psi^{(i)}_1(d)} \xrightarrow{p} \rho_R(d),$$

by noticing the term $\psi^{(i)}_1(d)$ is mean zero.

Therefore, by decomposing $\tilde{\rho}_R(d)$ in a similar manner as in C.29, we can show the
consistency of \( \hat{\rho}_R(d) \) to \( \rho_R(d) \). The proof for results (iv) is hence completed.

**Proof of results (v):**

This part can be easily shown using similar arguments as in the proof of results (iii). Recall the results in (C.20), we have \( \Sigma_R(d) \) is a \( s \times s \) asymptotic covariance matrix for term \( \psi_1^{(i)}(d) - \psi_0^{(i)}(d) \). Since \( \psi_1^{(i)}(d) - \psi_0^{(i)}(d) \) is equivalent to \( w_1^{(i)}(d) - w_0^{(i)}(d) \) where

\[
\frac{R_i}{r(X_i, d(X_i), M_i)} \frac{1}{A_i \pi(X_i) + (1 - A_i) \{1 - \pi(X_i)\}} + \theta \{X_i, d(X_i)\},
\]

and

\[
\frac{1 - R_i}{1 - r(X_i, d(X_i), M_i)} \frac{1}{A_i \pi(X_i) + (1 - A_i) \{1 - \pi(X_i)\}} + \theta \{X_i, d(X_i)\},
\]

we can show the theoretical form of \( \Sigma_R \) following the similar arguments in establishing (C.23) and (C.30) as

\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ w_1^{(i)}(d) - w_0^{(i)}(d) \right\} \left\{ w_1^{(i)}(d) - w_0^{(i)}(d) \right\}^\top \overset{p}{\to} \Sigma_R(d). \tag{C.32}
\]

Therefore, using similar arguments in proving (C.25), we can show

\[
\hat{\Sigma}_R(d) = \Sigma_R(d) + o_p(1).
\]

The proof of results (v) is hence completed.

**Proof of results (vi):**

We next show the consistency of the proposed calibrated value estimator \( \hat{V}(d) \) to the true value \( V(d) \). Recall (4.3) that

\[
\hat{V}(d) = \hat{V}_p(d) - \hat{\rho}(d)^\top \hat{\Sigma}^{-1}_M(d) \{ \hat{W}_p(d) - \hat{W}_U(d) \},
\]

where \( \hat{\rho}(d) \) is the estimator for \( \rho(d) \), and \( \hat{\Sigma}_M(d) \) is the estimator for \( \Sigma_M(d) \).

Based on the established results (ii) and (iii), with the assumption (A6) that covariates and outcomes are bounded, we have

\[
\hat{V}(d) = \hat{V}_p(d) - \rho(d)^\top \Sigma^{-1}_M(d) \{ \hat{W}_p(d) - \hat{W}_U(d) \} + o_p(1),
\]

\[
\hat{V}_p(d) = V(d) + o_p(1) \quad \text{and} \quad \hat{W}_p(d) - \hat{W}_U(d) = o_p(1)
\]

Accoding to (C.22) and (C.8), we have \( \hat{V}_p(d) = V(d) + o_p(1) \) and \( \hat{W}_p(d) - \hat{W}_U(d) = o_p(1) \).
under $X_p \sim X_U$. Hence, it is immediate that

$$\hat{V}(d) = V(d) + o_p(1) = V(d) + o_p(1).$$

The proof of results (vi) is hence completed.

**Proof of results (vii):**

Lastly, the consistency of the proposed calibrated value estimator $\hat{V}_R(d)$ to the true value $V(d)$ can be shown using the similar arguments in the proof of results (vi) by utilizing definition of $\hat{V}_R(d)$, the results (iv) and (v), and (C.22) and (C.20). We omit the details for brevity.

### C.4.6 Proof of Theorem 4.4.2

In this section, we prove the asymptotic normality of the proposed calibrated value estimator. The proof consists of three parts. In part 1, we aim to show

$$\hat{V}(\hat{d}) = V_n(\hat{d}) + o_p(N_p^{-1/2}), \quad (C.33)$$

where

$$V_n(\hat{d}) = \frac{1}{N_p} \sum_{i=1}^{N_p} v_p^{(i)}(\hat{d}) - \rho(\hat{d})^\top \Sigma_{\hat{M}}^{-1}(\hat{d}) \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} w_p^{(i)}(\hat{d}) - \frac{1}{N_U} \sum_{i=1}^{N_U} w_{ij}^{(i)}(\hat{d}) \right\}. $$

Then, in the second part, we establish

$$V_n(\hat{d}) = V_n(d^{opt}) + o_p(N_p^{-1/2}).$$

The above two steps yields that

$$\hat{V}(\hat{d}) = V_n(d^{opt}) + o_p(N_p^{-1/2}). \quad (C.34)$$

Lastly, based on (C.34), we show

$$\sqrt{N_p} \left\{ \hat{V}(\hat{d}) - V(d^{opt}) \right\} \sim N\left[0, \sigma^2(d^{opt}) - \rho(d^{opt})^\top \Sigma_{d^{opt}}^{-1}(d^{opt}) \rho(d^{opt}) \right].$$

**Proof of Part 1:** Based on (C.21), we have $\hat{V}_R(\hat{d}) = N_p^{-1} \sum_{i=1}^{N_p} v_p^{(i)}(\hat{d}) + o_p(N_p^{-1/2})$. Combining
this with the definition of $\hat{V}(\hat{d})$ yields that

$$\hat{V}(\hat{d}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \nu_p^{(i)}(\hat{d}) - \bar{\rho}(\hat{d})^\top \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2}).$$  \hfill (C.35)

Thus, to show (C.33), it is sufficient to show

$$\eta_1 = \rho(\hat{d})^\top \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2}).$$  \hfill (C.36)

According to (C.8), we have

$$\hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) = O_p(N_p^{-1/2}).$$

This together with results (ii) in Theorem 4.4.1 that $\hat{\rho}(\hat{d}) = \rho(\hat{d}) + o_p(1)$ yields

$$\eta_1 = \hat{\rho}(\hat{d})^\top \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \}
= \{ \rho(\hat{d})^\top + o_p(1) \} \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \}
= \rho(\hat{d})^\top \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(1) \hat{\Sigma}^{-1}(\hat{d}) O_p(N_p^{-1/2}).$$

Since $O_p(N_p^{-1/2}) o_p(1) = o_p(N_p^{-1/2})$ and $\hat{\Sigma}^{-1}$ is bounded, we have the above equation as

$$\eta_1 = \rho(\hat{d})^\top \hat{\Sigma}^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2}).$$  \hfill (C.37)

Similarly, combining results (iii) in Theorem 4.4.1 that $\hat{\Sigma}(\hat{d}) = \Sigma(\hat{d}) + o_p(1)$ and the condition that $\rho$ is bounded, we can further replace $\hat{\Sigma}^{-1}(\hat{d})$ in (C.37), which yields that

$$\eta_1 = \rho(\hat{d})^\top \Sigma^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2})$$  \hfill (C.38)

$$= \rho(\hat{d})^\top \{ \Sigma^{-1}(\hat{d}) + o_p(1) \} \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2})$$

$$= \rho(\hat{d})^\top \Sigma^{-1}(\hat{d}) \{ \hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d}) \} + o_p(N_p^{-1/2}).$$

Combining (C.38) with (C.4), since $\rho$ and $\Sigma^{-1}$ are bounded, with $0 < t < +\infty$, we can
show

\[ \eta_1 = \rho(\hat{d})^T \Sigma_M^{-1}(\hat{d}) [\hat{W}_p(\hat{d}) - \hat{W}_U(\hat{d})] + o_p(N_p^{-1/2}) \]

\[ = \rho(\hat{d})^T \Sigma_M^{-1}(\hat{d}) \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} \sum_{i=1}^{N_U} w_p^{(i)}(\hat{d}) - \frac{1}{N_U} \sum_{i=1}^{N_U} w_U^{(i)}(\hat{d}) + (1 - \sqrt{t})o_p(N_p^{-1/2}) \right\} + o_p(N_p^{-1/2}) \]

\[ = \rho(\hat{d})^T \Sigma_M^{-1}(\hat{d}) \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} w_p^{(i)}(\hat{d}) - \frac{1}{N_U} \sum_{i=1}^{N_U} w_U^{(i)}(\hat{d}) \right\} + o_p(N_p^{-1/2}). \]

Thus, (C.36) is proved. This together with (C.35) yields (C.33). The proof of Part 1 is hence completed.

**Proof of Part 2:** We next focus on proving \( V_n(\hat{d}) = V_n(d^{opt}) + o_p(N_p^{-1/2}) \). Define a class of function

\[
\mathcal{F}_d(X_p, A_p, M_p, Y_p) = \left[ v_p(d) - \rho(d)^T \Sigma_M^{-1}(d) \left\{ w_p(d) - \frac{1}{N_U} \sum_{j=1}^{N_U} w_U^{(j)}(d) \right\} \right. \\
\left. - v_p(d^{opt}) + \rho(d^{opt})^T \Sigma_M^{-1}(d^{opt}) \left\{ w_p(d^{opt}) - \frac{1}{N_U} \sum_{j=1}^{N_U} w_U^{(j)}(d^{opt}) \right\} : d(\cdot) \in \Pi \right].
\]

where recall

\[ v_p(d) := \frac{I\{d(X_p)\}[Y_p - \mu_p(X_p,d(X_p))]}{A_p \pi_p(X_p) + (1 - A_p)\{1 - \pi_p(X_p)\}} + \mu_p\{X_p,d(X_p)\}, \]

and

\[ w_p(d) := \frac{I\{d(X_p)\}[M_p - \theta(X_p,d(X_p))]}{A_p \pi_p(X_p) + (1 - A_p)\{1 - \pi_p(X_p)\}} + \theta\{X_p,d(X_p)\}, \]

and \( \Pi \) denotes the space of decision rules of interest such as \( \Pi_1 \) or \( \Pi_2 \).

Under the assumption (A5), we have the class of decision rules \( \Pi \) is a Vapnik-Chervonenkis (VC) class of functions. By the conclusion of Lemma 2.6.18 in Van Der Vaart and Wellner (1996), we know the indicator function of a VC class of functions is still VC class.

Furthermore, under assumptions (A5) and (A6), following results (iv) in Lemma A.1 of Rai (2018), it can be shown that \( v_p(d) \) and \( w_p(d) \) are continuous with respect to \( d \in \Pi \). Therefore, we have \( \mathcal{F}_d \) belongs to the VC class, and its entropy \( J(\mathcal{F}_d) \) is finite.
Define the supremum of the empirical process indexed by $\mathcal{F}_d$ as

$$||G_n||_{\mathcal{F}} \equiv \sup_{d \in \Pi} \frac{1}{\sqrt{N_p}} \sum_{i=1}^{N_p} \mathcal{F}_d(X_{P,i}, A_{P,i}, M_{P,i}, Y_{P,i}) - \mathbb{E}[\mathcal{F}_d(X_P, A_P, M_P, Y_P)]$$

$$= \sup_{d \in \Pi} \sqrt{N_p} \{V_n(d) - V_n(d^{\text{opt}}) - V(d) + V(d^{\text{opt}})\}.$$ 

By the assumption (A6), we have $\tilde{B} \equiv \max_{1 \leq i \leq n} \mathcal{F}_d(X_{P,i}, A_{P,i}, M_{P,i}, Y_{P,i}) < \infty$. Define the asymptotic variance $\sigma^2_n \equiv \sup_{d \in \Pi} \mathcal{F}_d^2, \text{where } P$ is the common distribution of $\{X_P, A_P, M_P, Y_P\}$. Based on the central limit theorem, we have

$$\sqrt{N_p} \left\{V_n(d) - V(d)\right\} \rightsquigarrow N\left\{0, \sigma_0^2(d)\right\}, \text{for all } d \in \Pi.$$

This implies $\sigma^2_n = O(N_p^{-1/2}).$

It follows from the maximal inequality developed in Corollary 5.1 of Chernozhukov et al. (2014) that there exist some constant $v_0 \geq 1$ and $\tilde{C} > 0$ such that

$$\mathbb{E}[||G_n||_{\mathcal{F}}] \lesssim \sqrt{v_0 \sigma_n^2 \log\{\tilde{C} \mathcal{J}(\mathcal{F}_d)/\sigma_n\}} + \frac{v_0 \tilde{B}^2}{\sqrt{N_p}} \log\{\tilde{C} \mathcal{J}(\mathcal{F}_d)/\sigma_n\}.$$

The above right-hand-side is upper bounded by

$$O(1)\sqrt{\frac{1}{N_p} \log(N_p^{1/4})},$$

where $O(1)$ denotes some universal constant.

Hence, we have

$$\sqrt{N_p} \{V_n(\hat{d}) - V_n(d^{\text{opt}}) - V(\hat{d}) + V(d^{\text{opt}})\} = o_p(1). \quad \text{(C.39)}$$

Under the margin condition (A8), following Theorem 2.3 in Kitagawa and Tetenov (2018), we have

$$V(\hat{d}) = V(d^{\text{opt}}) + o_p(N_p^{-1/2}). \quad \text{(C.40)}$$

Combining (C.39) with (C.40), we have

$$V_n(\hat{d}) = V_n(d^{\text{opt}}) + o_p(N_p^{-1/2}).$$
This together with (C.33) proves (C.34). Thus, we complete the proof of Part 2.

**Proof of Part 3:** By the conclusion of Part 2, we have

\[
\frac{1}{\sqrt{N_p}} \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} v_i(d_{opt}) - \rho(d_{opt})^\top \Sigma_M^{-1}(d_{opt}) \right\} = \frac{1}{\sqrt{N_p}} \left\{ \frac{1}{N_U} \sum_{j=1}^{N_U} w_j(d_{opt}) - \frac{1}{N_U} \sum_{j=1}^{N_U} w_j(d_{opt}) \right\},
\]

Thus, to prove the asymptotic normality of the proposed calibrated value estimator, it is sufficient to show the asymptotic normality of \( \frac{1}{\sqrt{N_p}} \left\{ V_n(d_{opt}) - V(d_{opt}) \right\} \) based on Slutsky’s theorem.

Noticing

\[
V_n(d_{opt}) = \frac{1}{N_p} \sum_{i=1}^{N_p} v_i(d_{opt}) - \rho(d_{opt})^\top \Sigma_M^{-1}(d_{opt}) \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} w_i(d_{opt}) - \frac{1}{N_U} \sum_{j=1}^{N_U} w_j(d_{opt}) \right\},
\]

where \( \rho(d_{opt})^\top \Sigma_M^{-1}(d_{opt}) \) is a fixed constant.

By the central limit theorem with \( T = \lim_{N_p \to +\infty} t < +\infty \), we have

\[
\sqrt{N_p} \left\{ V_n(d_{opt}) - V(d_{opt}) \right\} \sim N\left\{ 0, \sigma^2(d_{opt}) \right\}.
\]

Next, we give the explicit form of \( \sigma^2(d_{opt}) \). Notice that

\[
\frac{1}{\sqrt{N_p}} \left\{ V_n(d_{opt}) - V(d_{opt}) \right\} = \frac{1}{\sqrt{N_p}} \sum_{i=1}^{N_p} \left\{ v_i(d_{opt}) - V(d_{opt}) \right\}
\]

\[
- \frac{1}{\sqrt{N_p}} \sum_{i=1}^{N_p} \rho(d_{opt})^\top \Sigma_M^{-1}(d_{opt}) \left\{ w_i(d_{opt}) - W(d_{opt}) \right\}
\]

\[
+ \frac{1}{\sqrt{N_U}} \sum_{j=1}^{N_U} \rho(d_{opt})^\top \Sigma_M^{-1}(d_{opt}) \left\{ w_j(d_{opt}) - W(d_{opt}) \right\},
\]

Since the two samples (\( P \) and \( U \)) are independently collected from two different separate studies, we have the independent and identically distributed double robust terms
\{v_p^{(i)}(d^{opt}), w_p^{(i)}(d^{opt})\}_{1 \leq i \leq N_p}$ are independent of $\{w_U^{(i)}(d^{opt})\}_{1 \leq i \leq N_U}$. Hence, the variance of $\sqrt{N_p}\left[V_n(d^{opt}) - V(d^{opt})\right]$ is given by
\[
\sigma^2 = \lim_{N_p \to \infty} \{\text{Var}(\eta_2) + \text{Var}(\eta_3) + \text{Var}(\eta_4) - 2\text{cov}(\eta_2, \eta_3)\}.
\] (C.41)

Using similar arguments in proving results (ii) and (iii) in Theorem 4.4.1, we have
\[
\lim_{N_p \to \infty} \text{Var}(\eta_2) = \sigma_Y^2(d^{opt})
\]
\[
\lim_{N_p \to \infty} \text{Var}(\eta_3) = \rho(d^{opt})^\top \Sigma_M^{-1}(d^{opt}) \Sigma_p(d^{opt}) \Sigma_M^{-1}(d^{opt}) \rho(d^{opt})
\]
\[
\lim_{N_p \to \infty} \text{Var}(\eta_4) = \rho(d^{opt})^\top \Sigma_M^{-1}(d^{opt}) T \Sigma_U(d^{opt}) \Sigma_M^{-1}(d^{opt}) \rho(d^{opt})
\]
\[
\lim_{N_p \to \infty} \text{cov}(\eta_2, \eta_3) = \rho(d^{opt})^\top \Sigma_M^{-1}(d^{opt}) \rho(d^{opt}).
\] (C.42)

Combining (C.41) with (C.42), using the definition that $\Sigma_M(d) = \Sigma_P(d) + T\Sigma_U(d)$, we have
\[
\sigma^2 = \sigma_Y^2(d^{opt}) - \rho(d^{opt})^\top \Sigma_M^{-1}(d^{opt}) \rho(d^{opt}).
\]

Thus, we have
\[
\sqrt{N_p}\left[V_n(d^{opt}) - V(d^{opt})\right] \sim N\left[0, \sigma_Y^2(d^{opt}) - \rho(d^{opt})^\top \Sigma_M^{-1}(d^{opt}) \rho(d^{opt})\right].
\]

The proof is hence completed.

**C.4.7 Proof of Theorem 4.4.3**

The proof of Theorem 4.4.3 follows the proof of Theorem 4.4.2. With a similar manner, we can show the asymptotic normality of $\sqrt{N_p}\left[\widehat{V}_R(d^{opt}) - V(d^{opt})\right]$ by three parts specified in Section C.4.6. The only difference is to replace the previous value difference between two samples $\{[\widehat{W}_p(d) - \widehat{W}_U(d)]\}$ with the rebalanced value difference based on joint sample $\{[\widehat{W}_l(d) - \widehat{W}_0(d)]\}$ with a sample ratio, under the consistency results (vi), (v), and (vii) in Theorem 4.4.1. We omit the details for brevity.
D.1 Inverse Propensity-Score Weighted Estimator

D.1.1 IPW Estimator for the Long-term Outcome

According to Lemma 5.3.1 and the law of large number, the value function $V(\beta)$ can be consistently estimated by

$$V_n(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta)\} \mu_U(M_{E,i}, X_{E,i}) \pi(X_{E,i}) + (1 - A_{E,i}) \{1 - \pi(X_{E,i})\}.$$

We posit parametric models for $\pi(x) \equiv \pi(x; \gamma)$ and $\mu_U(m, x) \equiv \mu_U(m, x; \lambda)$ with the true model parameter $\gamma$ and $\lambda$, respectively. Then the above $V_n(\beta)$ can be rewritten as the model-based form,

$$V_n^*(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta)\} \mu_U(M_{E,i}, X_{E,i}; \lambda) \pi(X_{E,i}; \gamma) + (1 - A_{E,i}) \{1 - \pi(X_{E,i}; \gamma)\},$$

where $\pi(x; \gamma)$ can be estimated in the experimental sample, denoted as $\pi(x; \hat{\gamma})$, and $\mu_U(m, x; \lambda)$
can be estimated in the auxiliary sample, denoted as \( \mu_U(m, x; \lambda) \). Then, by replacing the implicit functions in \( V_n^*(\beta) \) with their parametric estimators, it is straightforward to give the following IPW estimator for the value function \( V(\beta) \),

\[
\widehat{V}(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}, \beta)] \mu_U(M_{E,i}, X_{E,i}; \hat{\lambda})}{A_{E,i} \pi(X_{E,i}; \hat{\gamma}) + (1 - A_{E,i})(1 - \pi(X_{E,i}; \hat{\gamma}))}.
\] (D.1)

Define \( \hat{\beta} = \arg \max_{\beta} \widehat{V}(\beta) \) with subject to \( ||\beta||_2 = 1 \) for identifiability purpose, with the corresponding estimated value function \( \widehat{V}(\hat{\beta}) \).

**D.1.2 Theoretical Results of the IPW Estimator**

First, we establish some theoretical results for the IPW estimator as a middle step to prove the results for the AIPW estimator. Here, we use \( \phi_X(X) \) and \( \phi_M(M) \) to represent appropriate basis functions for \( X \) and \( M \), respectively. The following theorem gives the consistency result of our IPW estimator for the value function to the true. The proof is provided in Section D.2.3.

**Theorem D.1.1.** *(Consistency)* When (A1)-(A9) and (A11) hold, given \( \forall \beta \), we have

\[
\widehat{V}(\beta) = V(\beta) + o_p(1).
\]

Next, we establish the asymptotic normality of \( \sqrt{N_E} \left( \widehat{V}(\beta) - V(\beta_0) \right) \) through the following lemma that states the estimator \( \hat{\beta} \) has a cubic rate towards the true \( \beta_0 \). The proof is provided in Section D.2.4.

**Lemma D.1.1.** Under (A1)-(A11), we have

\[
N_E^{1/3} ||\hat{\beta} - \beta_0||_2 = O_p(1),
\] (D.2)

where \( || \cdot ||_2 \) is the \( L_2 \) norm.

We next show the asymptotic distribution of \( \widehat{V}(\hat{\beta}) \) as follows. The proof is provided in Section D.2.5.

**Theorem D.1.2.** *(Asymptotic Distribution)* When (A1)-(A11) are satisfied, we have

\[
\sqrt{N_E} \left( \widehat{V}(\beta) - V(\beta_0) \right) \overset{D}{\rightarrow} N(0, \sigma_{IPW}^2),
\] (D.3)
where \( \sigma_{IPW}^2 = t \sigma_U^2 + \sigma_{E,I}^2 \), and \( \sigma_U^2 = \mathbb{E}[\{\xi_i^{(U)}\}^2] \) and \( \sigma_{E,I}^2 = \mathbb{E}[\{\xi_i^{(E,I)}\}^2] \).

Here, \( \xi_i^{(U)} \equiv G_2^T H_2^{-1} \left[ \phi_X(X_{U,i}) \right] \{Y_{U,i} - \mu_U(M_{U,i}, X_{U,i}, \lambda)\} \) is the I.I.D. term in the auxiliary sample, and \( \xi_i^{(E,I)} \equiv G_1^T H_1^{-1} \phi_X(X_{E,i}) \{A_{E,i} - \pi(X_{E,i}, \gamma)\} + \frac{\mathbb{I}(A_{E,i} = d(X_{E,i}; \beta_0)) \mu_U(M_{E,i}, X_{E,i})}{A_{E,i}\pi(X_{E,i}, \gamma) + (1 - A_{E,i})(1 - \pi(X_{E,i}, \gamma))} - V(\beta_0) \) is the I.I.D. term in the experimental sample.

### D.2 Technical Proofs

#### D.2.1 Proof of Lemma 5.3.1

For any decision rule \( d(\cdot; \beta) \), we will show that \( V_n(\beta) \) is a consistent estimator of the value function \( V(\beta) \) under (A1)-(A5).

(A.) First, we rewrite the expectation of the I.I.D. summation term of \( V_n(\beta) \) using the law of iterated expectation with (A4) and (A5).

(a1.) Taking the iterated expectation on \( \{X_E, M_E\} \), we have

\[
\mathbb{E}[V_n(\beta)] = \mathbb{E}\left[ \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \mu_U(M_E, X_E) \mid X_E, M_E \right] \right]
\]

\[
= \mathbb{E}\left[ \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \mu_U(M_E, X_E) \mid X_E, M_E \right] \right]
\]

\[
= \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \mu_U(M_E, X_E) \mid X_E, M_E \right].
\]

(a2.) By Corollary 5.2.1 that \( \mu_U(M_E, X_E) = \mathbb{E}[Y_E | X_E, M_E] = \mathbb{E}[Y_E | X_E, M_E] \), thus

\[
\mathbb{E}[V_n(\beta)] = \mathbb{E}\left[ \mathbb{E}[Y_E | X_E, M_E] \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \mid X_E, M_E \right] \right].
\]

(a3.) By (A5) that \( Y_E \perp A_E \mid X_E, M_E \), we have

\[
\mathbb{E}[V_n(\beta)] = \mathbb{E}\left[ \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} Y_E \mid X_E, M_E \right] \right].
\]

(a4.) From the inverse of the law of iterated expectation, then

\[
\mathbb{E}[V_n(\beta)] = \mathbb{E}\left[ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} Y_E \right].
\]

240
(B.) Next, we proof above $E\{V_n(\beta)\}$ is a consistent estimator of the value function $V(\beta)$ under (A1)-(A3).

(b1.) Taking the iterated expectation on $\{X_E\}$, we have

$$E\{V_n(\beta)\} = E\left[ E\left\{ \frac{\mathbb{I}\{A_E = d(X_E; \beta)\}}{A_E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} Y_E \bigg| X_E \right\} \right].$$

(b2.) Use the fact that $\mathbb{I}\{A_E = d(X_E; \beta)\} = A d(X_E; \beta) + (1 - A_E)(1 - d(X_E; \beta))$, thus

$$E\{V_n(\beta)\} = E\left[ E\left\{ \frac{A Y_E}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} d(X_E; \beta) \right. \left. + E\left\{ \frac{(1 - A) Y_E}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} \{1 - d(X_E; \beta)\} \right].$$

(b3.) By (A1) that $Y_E = A Y_E^*(1) + (1 - A) Y_E^*(0)$, and the fact $AA = A$,

$$E\{V_n(\beta)\} = E\left[ E\left\{ \frac{A\{A Y_E^*(1) + (1 - A) Y_E^*(0)\}}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} d(X_E; \beta) \right. \left. + E\left\{ \frac{(1 - A)\{A Y_E^*(1) + (1 - A) Y_E^*(0)\}}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} \{1 - d(X_E; \beta)\} \right]$$

$$= E\left[ E\left\{ \frac{A Y_E^*(1)}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} d(X_E; \beta) \right. \left. + E\left\{ \frac{(1 - A) Y_E^*(0)}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} \{1 - d(X_E; \beta)\} \right].$$

(b4.) Applying (A2) that $\{Y_E^*(0), Y_E^*(1)\} \perp A \mid X_E$, we have

$$E\{V_n(\beta)\} = \left.E\{Y_E^*(1)\} E\left\{ \frac{A}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} d(X_E; \beta) \right.$$  

$$+ E\{Y_E^*(0)\} E\left\{ \frac{(1 - A)}{A E \pi(X_E) + (1 - A_E)(1 - \pi(X_E))} \bigg| X_E \right\} \{1 - d(X_E; \beta)\} \right].$$

241
\[
E \left[ E \{ Y^*_E(1)|X_E \} \right] = E \left[ \frac{1}{\pi(X_E)} P(A_E = 1|X_E) d(X_E; \beta) \right. \\
+ \left. E \{ Y^*_E(0)|X_E \} \right] E \left[ \frac{1}{1 - \pi(X_E)} P(A_E = 0|X_E) \right] \{1 - d(X_E; \beta)\}]
\]

(b5.) Based on (A3) that \(0 < \pi(x) < 1\) for all \(x \in X_E\), as well as the inverse of the iterated expectation, we finally show that
\[
E \{ V_n(\beta) \} = E \left[ E \{ Y^*_E(1)|X_E \} d(X_E; \beta) + E \{ Y^*_E(0)|X_E \} \{1 - d(X_E; \beta)\} \right]
\]
\[
= E \{ Y^*_E(0)\{1 - d(X_E; \beta)\} + Y^*_E(1)d(X_E; \beta)\} = V(\beta). \quad \square
\]

D.2.2 Proof of Lemma 5.3.2

Lemma 5.3.2 can be easily shown through the technic of the law of iterated expectation with (A4) and (A5).

(A.) By taking the iterated expectation on \(\{M_E\}\), we have
\[
E_{Y_E|X_E} \{ Y_E|A_E = d(X_E; \beta), X_E \} = E_{M_E|X_E} \{ E_{Y_E|X_E} \{ Y_E|A_E = d(X_E; \beta), X_E, M_E \} \}. \]

(B.) By (A5) that \(Y_E \perp \perp A_E|X_E, M_E\), we have
\[
E_{M_E|X_E} \{ E_{Y_E|X_E} \{ Y_E|A_E = d(X_E; \beta), X_E, M_E \} \} = E_{M_E|X_E} \{ E_{Y_E|X_E} \{ Y_E|M_E, X_E \}|A_E = d(X_E; \beta), X_E \}. \]

(C.) By Corollary 5.2.1 that \(\mu_U(M_E, X_E) = E\{Y_U|X_E, M_E\} = E\{Y_E|X_E, M_E\}\), thus
\[
E_{M_E|X_E} \{ E_{Y_E} \{ Y_E|M_E, X_E \}|A_E = d(X_E; \beta), X_E \} = E_{M_E|X_E} \{ \mu_U(M_E, X_E)|A_E = d(X_E; \beta), X_E \}. \quad \square
\]

D.2.3 Proof of Theorem D.1.1

To show that \(\hat{V}(\beta) = V(\beta) + o_p(1)\), it is sufficient to show that \(V^*_n(\beta) = V(\beta) + o_p(1)\) and \(\hat{V}(\beta) = V^*_n(\beta) + O_p(N^{-\frac{1}{2}}_E)\).

242
(A.) First, given a decision rule \(d(\cdot; \beta)\), by the Weak Law of Large Number, we have

\[
V_n^*(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)]}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]} \mu_U(M_{E,i}, X_{E,i}; \lambda) 
\]

\[
\longrightarrow_p E \left[ \frac{\mathbb{I}[A_E = d(X_E; \beta)]}{A_E \pi(X_E; \gamma) + (1 - A_E)[1 - \pi(X_E; \gamma)]} \mu_U(M_E, X_E; \lambda) \right] = V(\beta).
\]

That is, \(V_n^*(\beta) = V(\beta) + o_p(1)\).

(B.) Next, we show the following is \(O_p(N_E^{-1})\),

\[
\tilde{V}(\beta) - V_n^*(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \left[ \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)]}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]} \mu_U(M_{E,i}, X_{E,i}; \lambda) \right] 
\]

\[
- \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)]}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]} \mu_U(M_{E,i}, X_{E,i}; \lambda) \right].
\]

(B1.) By (A8), with appropriate parametric model for \(\mu_U(m, x; \lambda)\), we can present the estimator \(\hat{\lambda}\) in the auxiliary sample as

\[
\sqrt{N_U}(\hat{\lambda} - \lambda) = H_2^{-1} \frac{1}{\sqrt{N_U}} \sum_{i=1}^{N_U} \left[ \frac{\phi_X(X_{U,i})}{\phi_M(M_{U,i})} \right] \{Y_{U,i} - \mu_U(M_{U,i}, X_{U,i}; \lambda)\} + o_p(1), \quad (D.5)
\]

where \(H_2 \equiv \lim_{N_U \to \infty} \frac{1}{N_U} \sum_{i=1}^{N_U} \left[ \frac{\phi_X(X_{U,i})}{\phi_M(M_{U,i})} \right] \{\partial \mu_U(M_{U,i}, X_{U,i}; \lambda) / \partial \lambda\}^\top\).

Similarly, according to (A11), for the IPW estimator, we can present the estimator \(\hat{\gamma}\) in the experimental sample as

\[
\sqrt{N_E}(\hat{\gamma} - \gamma) = H_1^{-1} \frac{1}{\sqrt{N_E}} \sum_{i=1}^{N_E} \phi_X(X_{E,i})\{A_{E,i} - \pi(X_{E,i}; \gamma)\} + o_p(1), \quad (D.6)
\]

where \(H_1 \equiv \lim_{N_E \to \infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \phi_X(X_{E,i})\{\partial \pi(X_{E,i}; \gamma) / \partial \gamma\}^\top\).
Take the Taylor Expansion on \( f_i(\tilde{\gamma}, \tilde{\lambda}; \beta) \) at \((\gamma, \lambda)\), we have

\[
f_i(\tilde{\gamma}, \tilde{\lambda}; \beta) - f_i(\gamma, \lambda; \beta) = \{\partial f_i(\tilde{\gamma}, \tilde{\lambda})/\partial \gamma\}(\tilde{\gamma} - \gamma) + \{\partial f_i(\tilde{\gamma}, \tilde{\lambda})/\partial \lambda\}(\tilde{\lambda} - \lambda)
\]

\[
= \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)] \mu_U(M_{E,i}, X_{E,i}; \lambda)}{[A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]]^2 (1 - 2A_{E,i})[\partial \pi(X_{E,i}; \gamma)/\partial \gamma](\tilde{\gamma} - \gamma)} (1 - 2A_{E,i})[\partial \pi(X_{E,i}; \gamma)/\partial \gamma](\tilde{\lambda} - \lambda) + o_p(N_E^{-1}).
\]

(b2.) Let

\[
\xi_{i,1} = \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)] \mu_U(M_{E,i}, X_{E,i}; \lambda)}{[A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]]^2 (1 - 2A_{E,i})[\partial \pi(X_{E,i}; \gamma)/\partial \gamma]},
\]

and

\[
\xi_{i,2} = \frac{\mathbb{I}[A_{E,i} = d(X_{E,i}; \beta)] \mu_U(M_{E,i}, X_{E,i}; \lambda)}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]}[\partial \mu_U(M_{E,i}, X_{E,i}; \lambda)/\partial \lambda],
\]

by (A3) that \(0 < \pi(x; \gamma) < 1\), and (A7) that \(\mu_U(m, x; \lambda), \partial \pi(x; \gamma)/\partial \gamma\) and \(\partial \mu_U(m, x; \lambda)/\partial \lambda\) are bounded, applying the Weak Law of Large Number, we have \(\frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,1} \overset{p}{\longrightarrow} B_1 < \infty\), and \(\frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,2} \overset{p}{\longrightarrow} B_2 < \infty\), as \(N_E \rightarrow \infty\).

(b3.) By rearranging the equations, we have

\[
\hat{V}(\beta) - V^*_n(\beta) = \frac{1}{N_E} \sum_{i=1}^{N_E} \{f_i(\tilde{\gamma}, \tilde{\lambda}; \beta) - f_i(\gamma, \lambda; \beta)\}
\]

\[
=(\tilde{\gamma} - \gamma) \frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,1} + (\tilde{\lambda} - \lambda) \frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,2} + o_p(N_E^{-1}),
\]

where \(\tilde{\gamma} - \gamma = O_E(N_E^{-\frac{1}{2}}), \tilde{\lambda} - \lambda = O_E(N_E^{-\frac{1}{2}}), \frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,1} = o_p(1) + B_1, \) and \(\frac{1}{N_E} \sum_{i=1}^{N_E} \xi_{i,2} = o_p(1) + B_2\).

By (A9) that \(t = \sqrt{\frac{N_E}{N_0}}\) with \(0 < t < +\infty\), thus,

\[
\hat{V}(\beta) - V^*_n(\beta) = O_E(N_E^{-\frac{1}{2}}).
\]

Therefore, \(\hat{V}(\beta) - V(\beta) = \hat{V}(\beta) - V^*_n(\beta) + V^*_n(\beta) - V(\beta) = O_E(N_E^{-\frac{1}{2}}) + o_p(1) = o_p(1).\)
D.2.4 Proof of Lemma D.1.1 and Lemma 5.3.3

(A.) First, we show that $\widehat{\beta}$ converges in probability to $\beta_0$ as $N_E \to \infty$, by checking three conditions of the Argmax Theorem:

(a1.) By (A10) that the true value function $V(\beta)$ has twice continuously differentiable at an inner point of maximum $\beta_0$.

(a2.) By the conclusion of Theorem D.1.1, $\widehat{V}(\beta) = V(\beta) + o_p(1)$, i.e., for $\forall \beta$

$$\widehat{V}(\beta) \xrightarrow{p} V(\beta), \quad \text{as} \quad N_E \to \infty.$$  

(a3.) Since $\widehat{\beta} = \arg\max_{\beta} \widehat{V}(\beta)$, we have the estimated ODR as $d(X_E; \widehat{\beta}) = I\{\phi_X(X_E)^\top \widehat{\beta} > 0\}$ and the corresponding value function $\widehat{V}(\widehat{\beta})$ such that $\widehat{V}(\widehat{\beta}) \geq \sup_{\beta \in B} \widehat{V}(\beta)$.

Thus, we have $\widehat{\beta} \xrightarrow{p} \beta_0$ as $N_E \to \infty$.

(B.) Next, we show that the convergence rate of $\widehat{\beta}$ is $N^{1/3}_E$, i.e. $N^{1/3}_E \|\widehat{\beta} - \beta_0\|_2 = O_p(1)$, where $\|\cdot\|_2$ is $L_2$ norm, via checking three conditions of the Theorem 14.4: Rate of convergence in Kosorok (2008):

(b1.) For every $\beta$ in a neighborhood of $\beta_0$, i.e. $\|\beta - \beta_0\|_2 < \delta$, by (A10), we take the second order Taylor expansion of $V(\beta)$ at $\beta = \beta_0$,

$$V(\beta) - V(\beta_0) = V'(\beta_0)\|\beta - \beta_0\|_2 + \frac{1}{2} V''(\beta_0)\|\beta - \beta_0\|_2^2 + o\{\|\beta - \beta_0\|_2^2\}$$

(by $V'(\beta_0) = 0$)$$= \frac{1}{2} V''(\beta_0)\|\beta - \beta_0\|_2^2 + o\{\|\beta - \beta_0\|_2^2\}.$$  

Since $V''(\beta_0) < 0$, there exist $c_1 = -\frac{1}{2} V''(\beta_0) > 0$ such that $V(\beta) - V(\beta_0) \leq c_1 \|\beta - \beta_0\|_2^2$ holds.

(b2.) For all $N_E$ large enough and sufficiently small $\delta$, the centered process $\widehat{V} - V$ satisfies

$$\text{E}^* \sup_{\|\beta - \beta_0\|_2 < \delta} \sqrt{N_E} |\widehat{V}(\beta) - V(\beta) - \{\widehat{V}(\beta_0) - V(\beta_0)\}|$$

$$= \text{E}^* \sup_{\|\beta - \beta_0\|_2 < \delta} \sqrt{N_E} |\widehat{V}(\beta) - V_n(\beta) + V_n(\beta) - V(\beta) - \{\widehat{V}(\beta_0) - V_n(\beta_0) + V_n(\beta_0) - V(\beta_0)\}|$$  

245
\[
\begin{align*}
\leq & E^* \sup_{\|\beta - \beta_0\|_2 < \delta} \sqrt{N_E} \left| \bar{V}(\beta) - V_n^*(\beta) - \{ \bar{V}(\beta_0) - V_n^*(\beta_0) \} \right| \\
+ & E^* \sup_{\|\beta - \beta_0\|_2 < \delta} \sqrt{N_E} \left| V_n^*(\beta) - V(\beta) - \{ V_n^*(\beta_0) - V(\beta_0) \} \right|
\end{align*}
\]

where \( E^* \) is the outer expectation.

We first derive two results (b2.1) and (b2.2) to bound \( \eta_1 \) and \( \eta_2 \), respectively, and then we are able to show the second condition of Theorem 14.4 is satisfied.

(b2.1) First, recalling the definition made in Equation (D.4) that

\[
f_i(\gamma, \lambda; \beta) = \frac{\mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta)\}}{A_{E,i}\pi(X_{E,i}; \gamma) + (1 - A_{E,i})\{1 - \pi(X_{E,i}; \gamma)\}} \mu_U(M_{E,i}, X_{E,i}; \lambda),
\]

with the fact that

\[
\begin{align*}
\mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta)\} - \mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta_0)\} \\
= A_{E,i}\mathbb{I}\{\phi_X(X_{E,i})^\top \beta > 0\} + (1 - A_{E,i})\left[1 - \mathbb{I}\{\phi_X(X_{E,i})^\top \beta > 0\}\right] \\
- A_{E,i}\mathbb{I}\{\phi_X(X_{E,i})^\top \beta_0 > 0\} - (1 - A_{E,i})\left[1 - \mathbb{I}\{\phi_X(X_{E,i})^\top \beta_0 > 0\}\right] \\
= (2A_{E,i} - 1)\left[\mathbb{I}\{\phi_X(X_{E,i})^\top \beta > 0\} - \mathbb{I}\{\phi_X(X_{E,i})^\top \beta_0 > 0\}\right].
\end{align*}
\]

we have,

\[
\begin{align*}
V_n^*(\beta) - V_n^*(\beta_0) &= \frac{1}{N_E} \sum_{i=1}^{N_E} \left\{ f_i(\gamma, \lambda; \beta) - f_i(\gamma, \lambda; \beta_0) \right\} \\
&= \frac{1}{N_E} \sum_{i=1}^{N_E} \mu_U(M_{E,i}, X_{E,i}; \lambda)\mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta)\} - \mathbb{I}\{A_{E,i} = d(X_{E,i}; \beta_0)\} \\
&= \frac{1}{N_E} \sum_{i=1}^{N_E} \mu_U(M_{E,i}, X_{E,i}; \lambda)(2A_{E,i} - 1)\left[\mathbb{I}\{\phi_X(X_{E,i})^\top \beta > 0\} - \mathbb{I}\{\phi_X(X_{E,i})^\top \beta_0 > 0\}\right].
\end{align*}
\]

Then, we define a class of function

\[
\mathcal{F}_\beta^1(x, a, m) = \left\{ \frac{\mu_U(m, x; \lambda)(2a - 1)\left[\mathbb{I}\{\phi_X(x)^\top \beta > 0\} - \mathbb{I}\{\phi_X(x)^\top \beta_0 > 0\}\right]}{a\pi(x; \gamma) + (1 - a)(1 - \pi(x; \gamma))} : \|\beta - \beta_0\|_2 < \delta \right\}.
\]

Let \( M_1 = \sup_{\|\beta - \beta_0\|_2 < \delta} \left| \frac{\mu_U(m, x; \lambda)(2a - 1)}{a\pi(x; \gamma) + (1 - a)(1 - \pi(x; \gamma))} \right| \), by (A7) that \( \mu_U(m, x; \lambda) \) and \( \pi(x; \gamma) \) both bounded, we have \( M_1 < \infty \). Then, we define the envelope of \( \mathcal{F}_\beta^1 \) as \( F_1 = M_1\mathbb{I}\{1 - \delta \leq \phi_X(x)^\top \beta_0 \leq 1 + \delta\} \);
by (A6) that the density function of covariate \( f_X(x) \) is bounded away from 0 and \( \infty \), thus,

\[
\| F_1 \|_{l_2} = M_1 \sqrt{P \{ 1 - \delta \leq \phi_X(x)^\top \beta_0 \leq 1 + \delta \}} = M_1 \sqrt{\phi_X(\beta_0) \cdot 2\delta} = M_1 \sqrt{2\phi_X(\beta_0)\delta^{\frac{1}{2}}} < \infty.
\]

Since \( \mathcal{F}_\beta^1 \) is an indicate function, by the conclusion of the Lemma 2.6.15 and Lemma 2.6.18 (iii) in Wellner et al. (2013), \( \mathcal{F}_\beta^1 \) is a VC (and hence Donsker) class of functions. Thus, the entropy of the class function \( \mathcal{F}_\beta^1 \) denoted as \( J_1^* (1, \mathcal{F}^1) \) is finite, i.e., \( J_1^* (1, \mathcal{F}^1) < \infty \).

Next, we consider the following empirical process indexed by \( \beta \),

\[
G_n \mathcal{F}_\beta^1 = \frac{1}{\sqrt{N_E}} \sum_{i=1}^{N_E} \left\{ \mathcal{F}_\beta^1 (X_{E,i}, A_{E,i}, M_{E,i}) - E \mathcal{F}_\beta^1 (X_{E,i}, A_i, M_{E,i}) \right\}.
\]

Note that \( G_n \mathcal{F}_\beta^1 = \sqrt{N_E} [ V_n^*(\beta) - V_n^*(\beta_0) - \{ V(\beta) - V(\beta_0) \} ] \) by Equation (D.9). Therefore, by applying Theorem 11.2 in Kosorok (2008), we have,

\[
\eta_2 = E^* \sup_{||\beta - \beta_0|| < \delta} \sqrt{N_E} \left| V_n^*(\beta) - V(\beta) - \{ V_n^*(\beta_0) - V(\beta_0) \} \right|
\]

\[
= E^* \sup_{||\beta - \beta_0|| < \delta} \left| G_n \mathcal{F}_\beta^1 \right| \leq c_1 J_1^* (1, \mathcal{F}^1) \| F_1 \|_{l_2} = c_1 J_1^* (1, \mathcal{F}^1) M_1 \sqrt{2\phi_X(\beta_0)\delta^{\frac{1}{2}}},
\]

where and \( c_1 \) is a finite constant.

Let \( C_1^* \equiv c_1 J_1^* (1, \mathcal{F}^1) M_1 \sqrt{2\phi_X(\beta_0)} \), since \( J_1^* (1, \mathcal{F}^1) \), \( M_1 \), and \( f_X(\cdot) \) are bounded, we have \( C_1^* < \infty \), i.e.,

\[
\eta_2 \leq C_1^* \delta^{\frac{1}{2}}.
\]  

(D.10)

(b2.2). First, we rewrite the form of \( \widehat{V}(\beta) - V_n^*(\beta) - \{ \widehat{V}(\beta_0) - V_n^*(\beta_0) \} \) by Equation (D.4) and Equation (D.8) as

\[
\widehat{V}(\beta) - V_n^*(\beta) - \{ \widehat{V}(\beta_0) - V_n^*(\beta_0) \} = \frac{1}{N_E} \sum_{i=1}^{N_E} \left\{ f_i(\widehat{\gamma}, \widehat{\lambda}; \beta) - f_i(\gamma, \lambda; \beta) - f_i(\widehat{\gamma}, \widehat{\lambda}; \beta_0) + f_i(\gamma, \lambda; \beta_0) \right\}
\]

\[
= \frac{1}{N_E} \sum_{i=1}^{N_E} \left\{ (2A_{E,i} - 1) \left[ \mathbb{I} (\phi_X(X_{E,i})^\top \beta > 0) - \mathbb{I} (\phi_X(X_{E,i})^\top \beta_0 > 0) \right] \right\}
\]

\[
\times \left[ \frac{\mu_U(M_{E,i}, X_{E,i}; \widehat{\lambda})}{A_{E,i} \pi(X_{E,i}; \widehat{\gamma}) + (1 - A_{E,i}) (1 - \pi(X_{E,i}; \widehat{\gamma}))} - \frac{\mu_U(M_{E,i}, X_{E,i}; \lambda)}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i}) (1 - \pi(X_{E,i}; \gamma))} \right].
\]

247
Based on (A8) and (A11), take the Taylor Expansion on above Equation at \((\gamma, \lambda)\), similar to Equation (D.7), we have

\[
\begin{align*}
\tilde{V}(\beta) - V^*_n(\beta) & - \{ \tilde{V}(\beta_0) - V^*_n(\beta_0) \} \\
= \frac{1}{N_E} \sum_{i=1}^{N_E} \left\{ \left[ 1 \{ \phi_X(X_{E,i}) \beta > 0 \} - 1 \{ \phi_X(X_{E,i}) \beta_0 > 0 \} \right] \\
\times \left[ - \frac{\mu_U(M_{E,i}, X_{E,i}; \lambda)(2A_{E,i} - 1)[\phi_2(\pi(X_{E,i}) \gamma)/\phi(\gamma)]}{A_{E,i} \pi(X_{E,i}; \gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i}; \gamma)]} \right] \right\}. \\
\end{align*}
\]

(D.11)

Next, we define two classes of function,

\[
\mathcal{F}_j(x, a, m) = \left\{ \frac{\mu_U(m, x; \lambda)(2a - 1)[\phi_2(\pi(x) \gamma)/\phi \gamma]}{a \pi(x; \gamma)} : ||\beta - \beta_0||_2 < \delta \right\}, \text{ and } \mathcal{F}_j(x, a, m) = \\
\left\{ \frac{(2a - 1)[\phi_2(a)(m, x; \lambda)/\phi \gamma]}{a \pi(x; \gamma)} : ||\beta - \beta_0||_2 < \delta \right\}.
\]

Let \(M_2 = \sup \left\{ \frac{\mu_U(m, x; \lambda)(2a - 1)[\phi_2(x) \gamma]/\phi(\gamma)}{a \pi(x; \gamma) + (1 - a)[1 - \pi(x; \gamma)]} \right\} \) and \(M_3 = \sup \left\{ \frac{(2a - 1)[\phi_2(m, x; \lambda)/\phi \gamma]}{a \pi(x; \gamma) + (1 - a)[1 - \pi(x; \gamma)]} \right\} \), then define the envelope of \(\mathcal{F}_j\) as \(F_j = M_j \mathcal{I}_j - \delta \leq \phi_X(x \gamma) \beta_0 \leq 1 + \delta\), for \(j = 2, 3\). Similarly to (b2.1), we have

\[
||F_j||_{p, 2} = M_j \sqrt{2f_X(\beta_0)} \delta^{\frac{1}{2}} < \infty,
\]

and \(\mathcal{F}_j\) is VC class of functions, thus, the entropy of the class function \(\mathcal{F}_j\) denoted as \(J^{*}_{\mathcal{F}}(1, \mathcal{F}_j)\) is finite, i.e., \(J^{*}_{\mathcal{F}}(1, \mathcal{F}_j) < \infty\), for \(j = 2, 3\).

Then, we construct two empirical processes indexed by \(\beta\),

\[
\mathbb{G}_n \mathcal{F}_j = \frac{1}{\sqrt{N_E}} \sum_{i=1}^{N_E} \left\{ \mathcal{F}_j(X_{E,i}, A_{E,i}, M_{E,i}) - \mathbb{E} \mathcal{F}_j(X_{E,i}, A_{i}, M_{E,i}) \right\}, \quad j = 2, 3.
\]

(D.12)

By Theorem 11.2 in Kosorok (2008), we have,

\[
\mathbb{E}^* \sup_{||\beta - \beta_0||_2 < \delta} ||\mathbb{G}_n \mathcal{F}_j|| \leq c_j J^{*}_{\mathcal{F}}(1, \mathcal{F}_j) ||F_j||_{p, 2} = c_j J^{*}_{\mathcal{F}}(1, \mathcal{F}_j) M_j \sqrt{2f_X(\beta_0)} \delta^{\frac{1}{2}}, \quad j = 2, 3,
\]

(D.13)

where \(c_2\) and \(c_3\) are finite constants, and let

\[
C_j^* \equiv c_j J^{*}_{\mathcal{F}}(1, \mathcal{F}_j) M_j \sqrt{2f_X(\beta_0)} < \infty.
\]

248
Finally, we rearrange the equations based on Equation (D.11) and Equation (D.12), so

\[ \eta_1 = E^* \sup_{||\beta - \beta_0|| < \delta} \sqrt{N_E} |\tilde{V}(\beta) - V_n^*(\beta) - \{\tilde{V}(\beta_0) - V_n^*(\beta_0)\}| \]

\[ = E^* \sup_{||\beta - \beta_0|| < \delta} \left| (\gamma - \gamma) G_n F^2_\beta + (\lambda - \lambda) G_n F^3_\beta + o_p(1) \right| \leq \frac{1}{\sqrt{N_E}} E^* \sup_{||\beta - \beta_0|| < \delta} \left| \sqrt{N_E} (\gamma - \gamma) G_n F^2_\beta \right| + \frac{1}{\sqrt{N_U}} E^* \sup_{||\beta - \beta_0|| < \delta} \left| \sqrt{N_U} (\lambda - \lambda) G_n F^3_\beta \right|. \]

By the H"{o}lder's Inequality, we have,

\[ \eta_1 \leq \frac{1}{\sqrt{N_E}} E \left| \sqrt{N_E} (\gamma - \gamma) \right| E^* \sup_{||\beta - \beta_0|| < \delta} \left| G_n F^2_\beta \right| + \frac{1}{\sqrt{N_U}} E \left| \sqrt{N_U} (\lambda - \lambda) \right| E^* \sup_{||\beta - \beta_0|| < \delta} \left| G_n F^3_\beta \right|. \]

Since \( \tilde{\gamma} - \gamma = O_E(N^{-1}_E) \) and \( \tilde{\lambda} - \lambda = O_E(N^{-1}_U) \), we have \( M_\gamma \equiv E \left| \sqrt{N_E} (\tilde{\gamma} - \gamma) \right| < \infty \) and \( M_\lambda \equiv E \left| \sqrt{N_U} (\tilde{\lambda} - \lambda) \right| < \infty \). By the results of (D.13) and (A9) that \( t = \frac{N_E}{N_U} \) with \( 0 < t < +\infty \), we have,

\[ \eta_1 \leq \frac{1}{\sqrt{N_E}} M_\gamma C^*_1 \delta^\frac{1}{2} + \frac{1}{\sqrt{N_U}} M_\lambda C^*_3 \delta^\frac{1}{2} = \frac{1}{\sqrt{N_E}} (M_\gamma C^*_2 + t M_\lambda C^*_3) \delta^\frac{1}{2}. \]  

(D.14)

By the results of (D.10) in (b2.1) and (D.14) in (b2.2), we have the centered process \( \tilde{V} - V \) satisfies

\[ E^* \sup_{||\beta - \beta_0|| < \delta} \sqrt{N_E} |\tilde{V}(\beta) - V(\beta) - \{\tilde{V}(\beta_0) - V(\beta_0)\}| \leq \eta_1 + \eta_2 \leq C^*_1 \delta^\frac{1}{2} + \frac{1}{\sqrt{N_E}} (M_\gamma C^*_2 + t M_\lambda C^*_3) \delta^\frac{1}{2}. \]

where \( M_\gamma, M_\lambda, \) and \( C^*_j, j = 1, 2, 3 \) are some finite constants. Let \( N_E \) goes infinite, we have

\[ E^* \sup_{||\beta - \beta_0|| < \delta} \sqrt{N_E} |\tilde{V}(\beta) - V(\beta) - \{\tilde{V}(\beta_0) - V(\beta_0)\}| \leq C^*_1 \delta^\frac{1}{2}. \]  

(D.15)

Let \( \phi_n(\delta) = C^*_1 \delta^\frac{1}{2} \), and \( \alpha = \frac{3}{2} < 2 \), check \( \frac{\phi_n(\delta)}{\delta^\alpha} = \frac{\delta^\frac{1}{2}}{\delta^\alpha} = \delta^{-1} \) is decreasing not depending on \( N_E \). Therefore, condition B holds.

(b3.) By \( \hat{\beta} \overset{p}{\to} \beta_0 \) as \( N_E \to \infty \) and \( \tilde{V}(\hat{\beta}) \geq \sup_{\beta \in B} \tilde{V}(\beta) \) shown previously, choose \( r_n = \)
\[ N_E^{1/3}, \text{ then } r_n \text{ satisfies} \]
\[ r_n^2 \phi_n(r_n^{-1}) = N_E^{2/3} \phi_n(N_E^{-1/3}) = N_E^{2/3} (N_E^{-1/3})^{1/2} = N_E^{2/3 - 1/6} = N_E^{1/2}. \]

Thus, condition C holds.

By the Theorem 14.4 in Kosorok (2008), we have \( N_E^{1/3} ||\hat{\beta} - \beta_0||_2 = O_p(1). \) □

Note that proof of Lemma 5.3.3 is just trivial extension of the above proof of Lemma D.1.1.

### D.2.5 Proof of Theorem D.1.2

To show the asymptotical distribution of the IPW estimator of the value function, we break down the following expression \( \sqrt{N_E} \{ \tilde{V}(\beta) - V(\beta_0) \} \) into two parts,

\[ \sqrt{N_E} \{ \tilde{V}(\beta) - V(\beta_0) \} = \sqrt{N_E} \{ \tilde{V}(\beta) - \tilde{V}(\beta_0) + \tilde{V}(\beta_0) - V(\beta_0) \} \]

\[ = \sqrt{N_E} \{ \tilde{V}(\beta) - \tilde{V}(\beta_0) \} + \sqrt{N_E} \{ \tilde{V}(\beta_0) - V(\beta_0) \}. \]

(A.) First, we show the first part

\[ \sqrt{N_E} \{ \tilde{V}(\beta) - \tilde{V}(\beta_0) \} = o_p(1), \]

which is sufficient to show \( \sqrt{N_E} \{ V(\beta) - V(\beta_0) \} = o_p(1) \) and \( \sqrt{N_E} \{ \tilde{V}(\beta) - \tilde{V}(\beta_0) \} - \{ V(\tilde{\beta}) - V(\beta_0) \} = o_p(1). \)

(a1.) First, by \( N_E^{1/3} ||\tilde{\beta} - \beta_0||_2 = O_p(1) \) and (A10), we take the second order Taylor expansion of \( V(\tilde{\beta}) \) at \( \beta_0 \), then

\[ \sqrt{N_E} \{ V(\tilde{\beta}) - V(\beta_0) \} = \sqrt{N_E} \left[ V'(\beta_0) ||\tilde{\beta} - \beta_0||_2 + \frac{1}{2} V''(\beta_0) ||\tilde{\beta} - \beta_0||_2^2 + o_p \{ ||\tilde{\beta} - \beta_0||_2 \} \right] \]

(by \( V'(\beta_0) = 0 \) = \( \sqrt{N_E} \left\{ \frac{1}{2} V''(\beta_0) O_E(N_E^{-\frac{3}{2}}) + o_p (N_E^{-\frac{1}{2}}) \right\} = \frac{1}{2} V''(\beta_0) O_E(N_E^{-\frac{3}{2}}) = o_p(1). \)

(D.16)

(a2.) Next, recall the result (D.15) in the proof of Lemma 5.3.3 that

\[ E^* \sup_{||\beta - \beta_0||_2 < \delta} \sqrt{N_E} |\tilde{V}(\beta) - V(\beta) - \{ \tilde{V}(\beta_0) - V(\beta_0) \} | \leq C_1^* \delta^{\frac{1}{2}}, \]

where \( C_1^* \) is a finite constant. Since \( ||\hat{\beta} - \beta_0||_2 = O_p(N_E^{-1/3}) \), i.e., \( ||\hat{\beta} - \beta_0||_2 = c_4 N_E^{-1/3}, \) where
\( c_4 \) is a finite constant, we have,
\[
\sqrt{N_E} \left[ \{ \tilde{V}(\beta) - \tilde{V}(\beta_0) \} - \{ V(\beta) - V(\beta_0) \} \right] \\
\leq E^* \sup_{\| \beta - \beta_0 \| < c_4 N_E^{-1/3}} \sqrt{N_E} \left| \tilde{V}(\beta) - V(\beta) - \{ \tilde{V}(\beta_0) - V(\beta_0) \} \right| \\
\leq C_1^* c_4 N_E^{-1/3} = C_1^* \sqrt{c_4} N_E^{-1/6} = o_p(1).
\]

(a3.) Thus, from the results of (D.16) and (D.17), we have,
\[
\sqrt{N_E} \left[ \tilde{V}(\beta) - \tilde{V}(\beta_0) \right] \\
= N_E \sum_{i=1}^{N_E} \frac{1}{N_E} \left[ \frac{1}{A_{E,i}} \frac{d(X_{E,i}; \beta_0)}{d \pi(X_{E,i}; \gamma)} \left( 1 - A_{E,i} \right) \left( 1 - \pi(X_{E,i}; \gamma) \right) \right] \frac{\partial \mu_U(M_{E,i}, X_{E,i}; \lambda)}{\partial \lambda} - \left( V(\beta) - V(\beta_0) \right) \\
= o_p(1) + o_p(1) = o_p(1).
\]

(B.) Next, we only need to show the asymptotical distribution of
\[
\sqrt{N_E} \left[ \tilde{V}(\beta_0) - V(\beta_0) \right] = \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{d(X_{E,i}; \beta_0)}{d \pi(X_{E,i}; \gamma)} \left( 1 - A_{E,i} \right) \left( 1 - \pi(X_{E,i}; \gamma) \right) \mu_U(M_{E,i}, X_{E,i}; \lambda) + o_p(N_E^{-1/2}), \\
\text{(D.18)}
\]

where \( G_1 \equiv \lim_{N_E \to +\infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{d(X_{E,i}; \beta_0)}{d \pi(X_{E,i}; \gamma)} \mu_U(M_{E,i}, X_{E,i}; \lambda) \),

and \( G_2 \equiv \lim_{N_E \to +\infty} \frac{1}{N_E} \sum_{i=1}^{N_E} \frac{d(X_{E,i}; \beta_0)}{d \pi(X_{E,i}; \gamma)} \left( 1 - A_{E,i} \right) \left( 1 - \pi(X_{E,i}; \gamma) \right) \frac{\partial \mu_U(M_{E,i}, X_{E,i}; \lambda)}{\partial \lambda} \).

(b2.) By Equation (D.6), Equation (D.5), and Equation (D.18), we have,
\[
\sqrt{N_E} \left[ \tilde{V}(\beta_0) - V(\beta_0) \right] \\
= \left( G_1 \right)^T \sqrt{N_E} (\gamma - \gamma) + G_2^T \left( \sqrt{N_U} (\hat{\lambda} - \lambda) + \sqrt{N_E} \left[ V_n^*(\beta_0) - V(\beta_0) \right] + o_p(1), \\
= \sqrt{N_E} \frac{1}{N_U} \varepsilon_{il} + \sqrt{N_E} \sum_{i=1}^{N_E} \varepsilon_{il}^{(E,i)} + o_p(1),
\]

251
where

\[ \xi_{i}^{(U)} \equiv G_{2}^{\top}H_{2}^{-1}\left[ \phi_{X}(X_{U,i})\phi_{M}(M_{U,i})\right] \{ Y_{U,i} - \mu_{U}(M_{U,i},X_{U,i};\lambda) \} \]

is the I.I.D. term in the auxiliary sample and

\[ \xi_{i}^{(E,1)} \equiv G_{1}^{\top}H_{1}^{-1}\phi_{X}(X_{E,i})[A_{E,i} - \pi(X_{E,i};\gamma)] + \frac{\mathbb{I}[A_{E,i} = d(X_{E,i};\beta_{0})]\mu_{U}(M_{E,i},X_{E,i};\tilde{\lambda})}{A_{E,i}\pi(X_{E,i};\gamma) + (1 - A_{E,i})[1 - \pi(X_{E,i};\gamma)]} - V(\beta_{0}) \]

is the I.I.D. term in the experimental sample.

(b3.) By (A9), we have \( t = \sqrt{\frac{N_{E}}{N_{U}}} \) and \( 0 < t < +\infty \). Since the estimation of \( \lambda \) is independent of the experimental sample, applying the central limit theorem, we have,

\[ \sqrt{N_{E}}(\tilde{V}(\beta_{0}) - V(\beta_{0})) \xrightarrow{\text{D}} N(0,\sigma_{IPW}^{2}) \]

where \( \sigma_{IPW}^{2} = t^{2}\sigma_{U}^{2} + \sigma_{E,1}^{2} \), and \( \sigma_{U}^{2} = \mathbb{E}[(\xi_{i}^{(U)})^{2}] \) and \( \sigma_{E,1}^{2} = \mathbb{E}[(\xi_{i}^{(E,1)})^{2}] \). \( \square \)

**D.2.6 Proof of Theorem 5.3.2**

Proof of Theorem 5.3.1 and Theorem 5.3.2 are trivial extensions of the proofs of Theorem D.1.1 and Theorem D.1.2. Here, we mainly address the augmented term of the AIPW estimator for the value function and show its asymptotic distribution.

(A.) Note that

\[
\tilde{V}_{AIP}(\beta) = \frac{1}{N_{E}} \sum_{i=1}^{N_{E}} \left[ \frac{\mathbb{I}[A_{E,i} = d(X_{E,i};\beta)]}{A_{E,i}\pi(X_{E,i};\tilde{\gamma}) + (1 - A_{E,i})[1 - \pi(X_{E,i};\tilde{\gamma})]} \mu_{U}(M_{E,i},X_{E,i};\tilde{\lambda}) + \left\{ 1 - \frac{\mathbb{I}[A_{E,i} = d(X_{E,i};\beta)]}{A_{E,i}\pi(X_{E,i};\tilde{\gamma}) + (1 - A_{E,i})[1 - \pi(X_{E,i};\tilde{\gamma})]} \mathbb{E}\left[ \mu_{U}(M_{E,i},X_{E,i};\tilde{\lambda})|A_{E,i} = d(X_{E,i};\beta),X_{E,i}\right] \right\} \right].
\]

Following the same procedure in (D.7), by taking the Taylor Expansion on \( \tilde{V}_{AIP}(\beta) \) at \((\gamma,\lambda)\), we have

\[
\tilde{V}(\beta_{0}) = G_{1}^{\top}(\tilde{\gamma} - \gamma) + G_{2}^{\top}(\tilde{\lambda} - \lambda) + G_{3}^{\top}(\tilde{\theta}_{0} - \theta_{0}) + G_{4}^{\top}(\tilde{\theta}_{1} - \theta_{1}) + V_{n,AIP}^{*}(\beta_{0}) + o_{p}(N_{E}^{-\frac{1}{2}}),
\]

where \( G_{3} \equiv \lim_{N_{E} \to +\infty} \frac{1}{N_{E}} \sum_{i=1}^{N_{E}} \frac{\mathbb{I}[A_{E,i} = d(X_{E,i};\beta_{0})]}{[A_{E,i}\pi(X_{E,i};\tilde{\gamma}) + (1 - A_{E,i})[1 - \pi(X_{E,i};\tilde{\gamma})]]^{2}} \frac{\partial}{\partial \gamma} \phi_{X}(X_{E,i}) \tilde{\gamma} \).
where $\lambda$ is the I.I.D. term in the experimental sample, and $G_4 \equiv \lim_{N_e \to +\infty} \frac{1}{N_e} \sum_{i=1}^{N_e} \left( 1 - A_E, i \right) \left( 1 - A_E, i \right) \{ \mu_U(M, E, i; X, E, i; \lambda) - \phi_X(X, E, i)^T \theta_0 \} + o_p(1),

\[ G_5 \equiv \lim_{N_e \to +\infty} \frac{1}{N_e} \sum_{i=1}^{N_e} \left( 1 - A_E, i \right) \left( 1 - A_E, i \right) \left( 1 - A_E, i \right) \{ \mu_U(M, E, i; X, E, i; \lambda) - \phi_X(X, E, i)^T \theta_0 \} + o_p(1), \]

where $\xi_i^E \equiv G_1 H_3^{-1} \phi_X(X, E, i)(1 - A_E, i) \{ \mu_U(M, E, i; X, E, i; \lambda) - \phi_X(X, E, i)^T \theta_0 \} + G_2 H_2^{-1} \phi_X(X, E, i) \{ \mu_U(M, E, i; X, E, i; \lambda) - \phi_X(X, E, i)^T \theta_0 \} + G_4 \left[ \frac{1}{N_E \sqrt{N_U}} \sum_{i=1}^{N_e} \xi_i^E + \frac{1}{N_E \sqrt{N_U}} \sum_{i=1}^{N_e} \xi_i^U \right] + o_p(1),

and $\xi_i^U \equiv G_2 H_2^{-1} \phi_X(X, U, i) \{ \mu_U(M, U, i; X, U, i; \lambda) - \phi_X(X, U, i)^T \theta_0 \} + o_p(1).$

By (A9), we have $t = \sqrt{\frac{N_E}{N_U}}$ and $0 < t < +\infty$. Since the estimation of $\lambda$ is independent of
the experimental sample, applying the central limit theorem, we have,

\[ \sqrt{N_E} \left( \hat{V}_{AIL}(\hat{\beta}^G) - V(\beta_0) \right) \Rightarrow N(0, \sigma^2_{AIL}) \]

where \( \sigma^2_{AIL} = t^2 \sigma^2_U + \sigma^2_E \), and \( \sigma^2_U = E[\{\epsilon(U)^2\}] \) and \( \sigma^2_E = E[\{\epsilon(E)^2\}] \). □

**D.3 Sensitivity Studies**

In this supplementary section, we investigate the finite sample performance of the proposed GEAR when the surrogacy assumption is violated in different extent, i.e. part of the information related to the long-term outcome cannot be collected or captured through intermediate outcomes. We consider the following Scenario 6 with \( r = 2 \) and \( s = 2 \).

**Scenario 6:**

\[
\begin{align*}
H^M(X) &= \begin{bmatrix} 0 \\ X^{(1)} \end{bmatrix}, \\
C^M(X) &= \begin{bmatrix} -0.5 + 0.4X^{(1)} - 0.6X^{(2)} \\ 0.5 + 0.6X^{(1)} - 0.4X^{(2)} \end{bmatrix}, \\
H^Y(X) &= X^{(2)}, \\
C^Y(X, M) &= M^{(1)} + M^{(2)},
\end{align*}
\]

where the true parameter of the ODR is \( \beta_0 = [0, 1/\sqrt{2}, -1/\sqrt{2}]^\top \) with the true value 0.333. We use the following \( M^{(1)}_{par} \) as one contaminated intermediate outcome we collected instead of the original \( M^{(1)} \),

\[ M^{(1)}_{par} = M^{(1)} + A(1-l)[-0.5 + 0.4X^{(1)}], \]

where the parameter \( l \) chosen from \( \{0, 0.2, 0.4, 0.6, 0.8, 1\} \) reflects the uncollected information related to the long-term outcome. When \( l = 1 \), we have the information of intermediate outcomes is fully collected. However, under \( l \in \{0, 0.2, 0.4, 0.6, 0.8\} \), the surrogacy assumption cannot hold anymore, since the long-term outcome is still dependent on the treatment given the information of \( M \) and \( X \).

Following the same estimation procedure as described in Section 5.4.1, we summarize the simulation results over 500 replications in Table D.1 for \( l = \{0, 0.4, 0.8\} \). Figure D.1 and Figure D.2 show how the bias of \( V(\hat{\beta}^G) \) towards the true value and the average rate of the correct decision made by the GEAR change as the parameter \( l \) (that indicates the
Table D.1: Empirical results of $\hat{\beta}^G$ and its estimated value $\widehat{\psi}_{API}(\hat{\beta}^G)$ under the GEAR for Scenario 6 when $l = \{0, 0.4, 0.8\}$. Note the true parameter of the ODR is $\beta_0 = [0, 1/\sqrt{2}, -1/\sqrt{2}]^T$ with the true value 0.333.

<table>
<thead>
<tr>
<th></th>
<th>$l = 0$</th>
<th></th>
<th></th>
<th></th>
<th>$l = 0.4$</th>
<th></th>
<th></th>
<th></th>
<th>$l = 0.8$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_E = 200$</td>
<td>400</td>
<td>800</td>
<td></td>
<td>$N_E = 200$</td>
<td>400</td>
<td>800</td>
<td></td>
<td>$N_E = 200$</td>
<td>400</td>
<td>800</td>
<td></td>
</tr>
<tr>
<td>$\widehat{\psi}_{API}(\hat{\beta}^G)$</td>
<td>0.546</td>
<td>0.494</td>
<td>0.470</td>
<td></td>
<td>0.505</td>
<td>0.472</td>
<td>0.434</td>
<td></td>
<td>0.470</td>
<td>0.434</td>
<td>0.401</td>
<td></td>
</tr>
<tr>
<td>SE[$\widehat{\psi}_{API}(\hat{\beta}^G)$]</td>
<td>0.154</td>
<td>0.113</td>
<td>0.091</td>
<td></td>
<td>0.156</td>
<td>0.113</td>
<td>0.086</td>
<td></td>
<td>0.156</td>
<td>0.118</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>E[$\widehat{\sigma}_{API}$]</td>
<td>0.158</td>
<td>0.118</td>
<td>0.092</td>
<td></td>
<td>0.160</td>
<td>0.120</td>
<td>0.093</td>
<td></td>
<td>0.162</td>
<td>0.121</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>$V(\hat{\beta}^G)$</td>
<td>0.265</td>
<td>0.276</td>
<td>0.284</td>
<td></td>
<td>0.285</td>
<td>0.296</td>
<td>0.298</td>
<td></td>
<td>0.293</td>
<td>0.306</td>
<td>0.314</td>
<td></td>
</tr>
<tr>
<td>Coverage prob. (%)</td>
<td>73.8</td>
<td>72.8</td>
<td>69.4</td>
<td></td>
<td>82.4</td>
<td>81.8</td>
<td>81.8</td>
<td></td>
<td>86.6</td>
<td>86.6</td>
<td>90.8</td>
<td></td>
</tr>
<tr>
<td>Correct Rate (%)</td>
<td>79.5</td>
<td>80.2</td>
<td>81.5</td>
<td></td>
<td>83.0</td>
<td>84.9</td>
<td>85.2</td>
<td></td>
<td>84.7</td>
<td>87.0</td>
<td>89.3</td>
<td></td>
</tr>
<tr>
<td>$\ell_2$ loss of $\hat{\beta}^G$</td>
<td>0.457</td>
<td>0.413</td>
<td>0.371</td>
<td></td>
<td>0.388</td>
<td>0.322</td>
<td>0.306</td>
<td></td>
<td>0.358</td>
<td>0.288</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(1)}$</td>
<td>-0.284</td>
<td>-0.292</td>
<td>-0.276</td>
<td></td>
<td>-0.184</td>
<td>-0.174</td>
<td>-0.191</td>
<td></td>
<td>-0.041</td>
<td>-0.059</td>
<td>-0.054</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0.790</td>
<td>0.794</td>
<td>0.785</td>
<td></td>
<td>0.765</td>
<td>0.765</td>
<td>0.775</td>
<td></td>
<td>0.730</td>
<td>0.736</td>
<td>0.739</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>-0.544</td>
<td>-0.534</td>
<td>-0.554</td>
<td></td>
<td>-0.618</td>
<td>-0.621</td>
<td>-0.602</td>
<td></td>
<td>-0.682</td>
<td>-0.674</td>
<td>-0.671</td>
<td></td>
</tr>
</tbody>
</table>

uncollected information of intermediate outcomes) changes, respectively.

Based on the results, our proposed method still has a reasonable performance when the surrogacy assumption is mildly violated. Specifically, the proposed GEAR achieves $V(\hat{\beta}^G) = 0.314$ in Scenario 6 ($V(\beta_0) = 0.333$) with an empirical coverage probability as 90.8% under $l = 0.8$ and $N_E = 800$. In addition, it is clear that including more intermediate outcomes that are highly correlated to the long-term outcome, could help to explain the treatment effect on the long-term outcome according to Figure D.1 and Figure D.2.
Figure D.1: The trend of the bias of $V(\hat{\beta}^G)$ under the GEAR over the parameter $l$.

Figure D.2: The trend of the average rate of the correct decision made by the GEAR over the parameter $l$. 