7-2020

Three essays on econometrics

Xin ZHENG
Singapore Management University, xin.zheng.2015@phdecons.smu.edu.sg

SINGAPORE MANAGEMENT UNIVERSITY

PHD DISSERTATION

Three Essays on Econometrics

Xin Zheng

supervised by
Professor LIANGJUN SU

July 2020

Three Essays on Econometrics

Xin Zheng

A DISSERTATION
In
ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment of the Requirements for the Degree of PhD in Economics

2020

Supervisor of Dissertation

PhD in Economics, Programme Director

Dissertation Committee:

Liangjun Su (Supervisor/Chair)
Lee Kong Chian Professor of Economics
School of Economics
Singapore Management University

Yichong Zhang
Assistant Professor of Economics
School of Economics
Singapore Management University

Peter C.B.Phillips
Yale University: Sterling Professor of Economics & Professor of Statistics
University of Auckland: Distinguished Professor
University of Southampton: Adjunct Professor of Economics
Singapore Management University: Distinguished Term Professor of Economics

Qu Feng
Associate Professor of Economics
School of Social Science
Nanyang Technological University

I hereby declare that this PhD dissertation is my original work.
I have duly acknowledged all the sources of information which have been used in this dissertation.

This dissertation has also not been submitted for any degree in any university previously.

Abstract

The dissertation includes three chapters on econometrics. The first chapter is about treatment effects and its application in randomized control trial. The second chapter is about specification test. The third chapter is about panel data model with fixed effects.

In the first chapter, we study the estimation and inference of the quantile treatment effect under covariate-adaptive randomization. We propose two estimation methods: (1) the simple quantile regression and (2) the inverse propensity score weighted quantile regression. For the two estimators, we derive their asymptotic distributions uniformly over a compact set of quantile indexes, and show that, when the treatment assignment rule does not achieve strong balance, the inverse propensity score weighted estimator has a smaller asymptotic variance than the simple quantile regression estimator. For the inference of method (1), we show that the Wald test using a weighted bootstrap standard error under-rejects. But for method (2), its asymptotic size equals the nominal level. We also show that, for both methods, the asymptotic size of the Wald test using a covariate-adaptive bootstrap standard error equals the nominal level. We illustrate the finite sample performance of the new estimation and inference methods using both simulated and real datasets.

In the second chapter, we propose a novel consistent model specification test based on the martingale difference divergence (MDD) of the error term given the covariates. The MDD equals zero if and only if error term is conditionally mean independent of the covariates. Our MDD test does not require any nonparametric estimation under the null or alternative and it is applicable even if we have many covariates in the regression model. We have established the asymptotic distributions of our test statistic under the null and under a sequence of Pitman local alternatives converging to the null at the usual parametric rate. We have conducted simulations to evaluate the finite sample performance of our test and compare it with its competitors. We find that our MDD test has superb performance in terms of both size and power and it generally dominates its competitors. In particular, it's the only test that has well controlled size in the presence of many covariates and reasonable power against high frequent alternatives as well. We apply our test to test for the correct specification of functional forms in gravity equations for four datasets. For all the datasets, we reject the log and level model coherently at 10% significance level. However, its competitors show mixed testing results for different datasets. The findings reveal the advantages of our test.

In the third chapter, we consider the Nickell bias problem in dynamic fixed effects multi-level panel data models with various kinds of multi-way error components. For some specifications of error components, there exist many different forms of within estimators which are shown to be of possibly different asymptotic properties. The forms of the estimators in our framework are given explicitly. We apply the split-sample jackknife approach to eliminate the bias. In practice, our results can be easily extended to multilevel panel data models with higher dimensions.

Contents

Acknowledgments

First, I want to thank my supervisor Liangjun Su for his kind advice and support. He is a perfect role model that I want to chase in the rest of my life. Second, I would like to thank Yichong Zhang for his profession, kindness and patience. Third, I want to thank Peter Phillips for his excellent lectures on time series analysis. Fourth, I would like to thank my family for the selfless support. I also want to thank all the members in our econometrics reading group and my classmates.

Chapter 1

Quantile Treatment Effects and Bootstrap Inference under Covariate-Adaptive Randomization

[1]

1.1 Introduction

The randomized control trial (RCT), as pointed out by Angrist and Pischke (2008), is one of the five most common methods (along with instrumental variable regressions, matching estimations, differences-in-differences, and regression discontinuity designs) for causal inference. Researchers can use the RCT to estimate not only average treatment effects (ATEs) but also quantile treatment effects (QTEs), which capture the heterogeneity of the sign and magnitude of treatment effects, varying depending on their place in the overall distribution of outcomes. For example, Muralidharan and Sundararaman (2011) estimate the QTE of teacher performance pay program on student learning via the difference of empirical quantiles of test scores between treatment and control groups. Duflo et al. (2013) and Banerjee et al. (2015) estimate the QTEs of audits on endline pollution and a group-lending microcredit program on informal borrowing, respectively, via linear quantile regressions (QRs). Crépon et al. (2015) estimate the QTE of microcredit on various household outcomes via a minimum distance method. Byrne et al. (2018) estimate the QTE of being informed on energy use via the inverse propensity score weighted (IPW) QR. Except Crépon et al. (2015), the other four pa-

[1]This is a co-authered work with Yichong Zhang.

pers all use the bootstrap to construct confidence intervals for their QTE estimates. However, RCTs have also been routinely implemented with covariate-adaptive randomization. Individuals are first stratified based on some baseline covariates, and then, within each stratum, the treatment status is assigned (independent of covariates) to achieve some balance between the sizes of treatment and control groups; as examples, see Imbens and Rubin (2015, Chapter 9) for a textbook treatment of the topic, and Duflo et al. (2007) and Bruhn and McKenzie (2009) for two excellent surveys on implementing RCTs in development economics. To achieve such balance, treatment status for different individuals usually exhibits a (negative) cross-sectional *dependence*. The standard inference procedures that rely on cross-sectional *independence* are usually conservative and lacking power. How do we consistently estimate QTEs under covariate-adaptive randomization? What are the asymptotic distributions for the QTE estimators, and how do we conduct proper bootstrap inference? These questions are as yet unaddressed.

We propose two ways to estimate QTEs: (1) the simple quantile regression (SQR) and (2) the IPW QR. We establish the weak limits for both estimators uniformly over a compact set of quantile indexes and show that the IPW estimator has a smaller asymptotic variance than the SQR estimator when the treatment assignment rule does not achieve strong balance.[2] If strong balance is achieved, then the two estimators are asymptotically first-order equivalent. For inference, we show that the Wald test combined with weighted bootstrap based critical values can lead to under-rejection for method (1), but its asymptotic size equals the nominal level for method (2). We also study the covariate-adaptive bootstrap which respects the cross-sectional dependence when generating the bootstrap sample. The estimator based on the covariate-adaptive bootstrap sample can mimic that of the original sample in terms of the standard error. Thus, using proper covariate-adaptive bootstrap based critical values, the asymptotic size of the Wald test equals the nominal level for both estimators.

As originally proposed by Doksum (1974), the QTE, for a fixed quantile index, corresponds to the horizontal difference between the marginal distributions of the potential outcomes for treatment and control groups. Firpo (2007) studies the identification and estimation of QTE under unconfoundedness. Our estimators (1) and (2) directly follow those in Doksum (1974) and Firpo (2007), respectively.

Shao et al. (2010) first point out that, under covariate-adaptive randomization, the usual two-sample t-test for the ATE is conservative. They then propose a covariate-adaptive bootstrap which can produce the correct standard error. Shao and Yu (2013) extend the results to generalized linear models. However, both groups of researchers parametrize the (transformed)

[2] We will define "strong balance" in Section 1.2.

conditional mean equation by a specific linear model and focus on a specific randomization scheme (covariate-adaptive biased coin method). Ma et al. (2018) derive the theoretical properties of ATE estimators based on general covariate-adaptive randomization under the linear model framework. Bugni et al. (2018) substantially generalize the framework to a fully non-parametric setting with a general class of randomization schemes. However, they mainly focus on the ATE and show that the standard two-sample t-test and the t-test based on the linear regression with strata fixed effects are conservative. They then obtain analytical estimators for the correct standard errors and study the validity of permutation tests. Hahn et al. (2011) study the IPW estimator for the ATE under adaptive randomization. However, they assume the treatment status is assigned completely independently across individuals. More recently, Bugni et al. (2019) study the estimation of ATE with multiple treatments and propose a fully saturated estimator. Tabord-Meehan (2018) study the estimation of ATE under an adaptive randomization procedure.

Our paper complements the above papers in four aspects. First, we consider the estimation and inference of the QTE, which is a function of quantile index τ. We rely on the empirical processes theories developed by van der Vaart and Wellner (1996) and Chernozhukov et al. (2014) to obtain uniformly weak convergence of our estimators over a compact set of τ. Based on the uniform convergence, we can construct not only point-wise but also uniform confidence bands. Second, we study the asymptotic properties of the IPW estimator under covariate-adaptive randomization. When the treatment assignment rule does not achieve strong balance, the IPW estimator is more efficient than the SQR estimator. Third, we investigate the weighted bootstrap approximation to the asymptotic distributions of the SQR and IPW estimators. We show that the weighted bootstrap ignores the (negative) cross-sectional dependence due to the covariate-adaptive randomization and over-estimates the asymptotic variance for the SQR estimator. However, the asymptotic variance for the IPW estimator does not rely on the randomization scheme implemented. Thus, the asymptotic size of the Wald test using the IPW estimator paired with the weighted bootstrap based critical values equals the nominal level. Fourth, we investigate the covariate-adaptive bootstrap approximation to the asymptotic distributions of the SQR and IPW estimators. We establish that, using either estimator paired with its corresponding covariate-adaptive bootstrap based critical values, the asymptotic size of the Wald test equals the nominal level. Shao et al. (2010) first propose the covariate-adaptive bootstrap and establish its validity for the ATE in a linear regression model under the null hypothesis that the treatment effect is not only zero but also homogeneous.[3] We modify the covariate-adaptive bootstrap and establish its validity

[3] We say the average treatment effect is homogeneous if the conditional average treatment effect given

for the QTE in the nonparametric setting proposed by Bugni et al. (2018). In addition, our results do not rely on the homogeneity of the treatment effect. Compared with the analytical inference, the two bootstrap inferences for QTEs we study in this paper avoid estimating the infinite-dimensional nuisance parameters such as the densities of the potential outcomes, and thus, the choices of tuning parameters. In addition, unlike the permutation tests studied in Bugni et al. (2018), the validity of bootstrap inferences does not require either strong balance condition or studentization. In particular, such studentization is cumbersome in the QTE context.

As the asymptotic variance for the IPW estimator does not depend on the treatment assignment rule implemented in RCTs, this estimator (and equivalently, the fully saturated estimator for the ATE) is suitable for settings where the knowledge of the exact treatment assignment rule is not available. Such scenario occurs when researchers are using an experiment that was run in the past and the randomization procedure may not have been fully described. It also occurs in subsample analysis, where sub-groups are defined using variables that may have not been used to form the strata and the treatment assignment rule for each sub-group becomes unknown. We illustrate this fact in the subsample analysis of the empirical application in Section 1.8.

The rest of the paper is organized as follows. In Section 1.2, we describe the model setup and notation. In Sections 1.3.1 and 1.3.2, we discuss the asymptotic properties of estimators (1) and (2), respectively. In Sections 1.4 and 1.5, we investigate the weighted and covariate-adaptive bootstrap approximations to the asymptotic distributions of estimators (1) and (2), respectively. In Section 1.6, we examine the finite-sample performance of the estimation and inference methods. In Section 1.7, we provide recommendations for practitioners. In Section 1.8, we apply the new methods to estimate and infer the average and quantile treatment effects of iron efficiency on educational attainment. In Section 1.9, we conclude. We provide proofs for all results in an appendix. We study the strata fixed effects quantile regression estimator and provide additional simulation results in the second online supplement.

1.2 Setup and Notation

First, denote the potential outcomes for treated and control groups as $Y(1)$ and $Y(0)$, respectively. The treatment status is denoted as A, where $A = 1$ means treated and $A = 0$ means untreated. The researcher can only observe $\{Y_i, Z_i, A_i\}_{i=1}^n$ where $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$, and Z_i is a collection of baseline covariates. Strata are constructed from Z using a function

covariates is the same as the unconditional one.

$S : \mathrm{Supp}(Z) \mapsto \mathcal{S}$, where $\mathcal{S}$ is a finite set. For $1 \le i \le n$, let $S_i = S(Z_i)$ and $p(s) = \mathbb{P}(S_i = s)$. Throughout the paper, we maintain the assumption that $p(s)$ is fixed w.r.t. n and is positive for every $s \in \mathcal{S}$.[4] We make the following assumption for the data generating process (DGP) and the treatment assignment rule.

Assumption 1. *(i)* $\{Y_i(1), Y_i(0), S_i\}_{i=1}^n$ *is i.i.d.*

(ii) $\{Y_i(1), Y_i(0)\}_{i=1}^n \perp\!\!\!\perp \{A_i\}_{i=1}^n | \{S_i\}_{i=1}^n$.

(iii) $\left\{ \left\{ \frac{D_n(s)}{\sqrt{n}} \right\}_{s \in \mathcal{S}} \middle| \{S_i\}_{i=1}^n \right\} \rightsquigarrow N(0, \Sigma_D)$ *a.s., where*

$$D_n(s) = \sum_{i=1}^n (A_i - \pi) 1\{S_i = s\} \quad \text{and} \quad \Sigma_D = \text{diag}\{p(s)\gamma(s) : s \in \mathcal{S}\}$$

with $0 \le \gamma(s) \le \pi(1 - \pi)$.

(iv) $\frac{D_n(s)}{n(s)} = o_p(1)$ *for* $s \in \mathcal{S}$*, where* $n(s) = \sum_{i=1}^n 1\{S_i = s\}$.

Several remarks are in order. First, Assumptions 1(i)–1(iii) are exactly the same as Bugni et al. (2018, Assumption 2.2). We refer interested readers to Bugni et al. (2018) for more discussion of these assumptions. Second, note that in Assumption 1(iii) the parameter π is the target proportion of treatment for each stratum and $D_n(s)$ measures the imbalance. Bugni et al. (2019) study the more general case that π can take distinct values for different strata. Third, we follow the terminology in Bugni et al. (2018), which follows that of Efron (1971) and Hu and Hu (2012), saying a treatment assignment rule achieves strong balance if $\gamma(s) = 0$. Fourth, we do not require that the treatment status is assigned independently. Instead, we only require Assumption 1(iii) or Assumption 1(iv), which condition is satisfied by several treatment assignment rules such as simple random sampling (SRS), biased-coin design (BCD), adaptive biased-coin design (WEI), and stratified block randomization (SBR). Bugni et al. (2018, Section 3) provide an excellent summary of these four examples. For completeness, we briefly repeat their descriptions below. Note that both BCD and SBR assignment rules achieve strong balance. Last, as $p(s) > 0$, Assumption 1(iii) implies Assumption 1(iv).

Example 1 (SRS)**.** *Let* $\{A_i\}_{i=1}^n$ *be drawn independently across* i *and of* $\{S_i\}_{i=1}^n$ *as Bernoulli random variables with success rate* π*, i.e., for* $k = 1, \cdots, n$*,*

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i=1}^n, \{A_j\}_{j=1}^{k-1}\right) = \mathbb{P}(A_k = 1) = \pi.$$

[4] We can also allow for the DGP to depend on n so that $p_n(s) = \mathbb{P}_n(S_i = s)$ and $p(s) = \lim p_n(s)$. All the results in this paper still hold as long as $n(s) \to \infty$ a.s. Interested readers can refer to the previous version of this paper on arXiv for more detail.

Then, Assumption 1(iii) holds with $\gamma(s) = \pi(1-\pi)$.

Example 2 (WEI)**.** *The design is first proposed by Wei (1978). Let $n_{k-1}(S_k) = \sum_{i=1}^{k-1} 1\{S_i = S_k\}$, $D_{k-1}(s) = \sum_{i=1}^{k-1}\left(A_i - \frac{1}{2}\right) 1\{S_i = s\}$, and*

$$\mathbb{P}\left(A_k = 1 \middle| \{S_i\}_{i=1}^k, \{A_i\}_{i=1}^{k-1}\right) = \phi\left(\frac{D_{k-1}(S_k)}{n_{k-1}(S_k)}\right),$$

where $\phi(\cdot) : [-1,1] \mapsto [0,1]$ is a pre-specified non-increasing function satisfying $\phi(-x) = 1 - \phi(x)$. Here, $\frac{D_0(S_1)}{0}$ is understood to be zero. Then, Bugni et al. (2018) show that Assumption 1(iii) holds with $\pi = \frac{1}{2}$ and $\gamma(s) = \frac{1}{4}(1 - 4\phi'(0))^{-1}$.

Example 3 (BCD)**.** *The treatment status is determined sequentially for $1 \leq k \leq n$ as*

$$\mathbb{P}\left(A_k = 1 | \{S_i\}_{i=1}^k, \{A_i\}_{i=1}^{k-1}\right) = \begin{cases} \frac{1}{2} & \text{if } D_{k-1}(S_k) = 0 \\ \lambda & \text{if } D_{k-1}(S_k) < 0 \\ 1-\lambda & \text{if } D_{k-1}(S_k) > 0, \end{cases}$$

where $D_{k-1}(s)$ is defined as above and $\frac{1}{2} < \lambda \leq 1$. Then, Bugni et al. (2018) show that Assumption 1(iii) holds with $\pi = \frac{1}{2}$ and $\gamma(s) = 0$.

Example 4 (SBR)**.** *For each stratum, $\lfloor \pi n(s) \rfloor$ units are assigned to treatment and the rest is assigned to control. Bugni et al. (2018) then show that Assumption 1(iii) holds with $\gamma(s) = 0$.*

Our parameter of interest is the τ-th QTE defined as

$$q(\tau) = q_1(\tau) - q_0(\tau),$$

where $\tau \in (0,1)$ is a quantile index and $q_j(\tau)$ is the τ-th quantile of random variable $Y(j)$ for $j = 0, 1$. For inference, although we mainly focus on the Wald test for the null hypothesis that $q(\tau)$ equals some particular value, our method can also be used to test hypotheses involving multiple or even a continuum of quantile indexes. The following regularity conditions are common in the literature of quantile estimations.

Assumption 2. *For $j = 0, 1$, denote $f_j(\cdot)$ and $f_j(\cdot|s)$ as the PDFs of $Y_i(j)$ and $Y_i(j)|S_i = s$, respectively.*

(i) $f_j(q_j(\tau))$ and $f_j(q_j(\tau)|s)$ are bounded and bounded away from zero uniformly over $\tau \in \Upsilon$ and $s \in \mathcal{S}$, where Υ is a compact subset of $(0,1)$.

(ii) $f_j(\cdot)$ and $f_j(\cdot|s)$ are Lipschitz over $\{q_j(\tau) : \tau \in \Upsilon\}$.

1.3 Estimation

1.3.1 Simple Quantile Regression

In this section, we propose to estimate $q(\tau)$ by a QR of Y_i on A_i. Denote $\beta(\tau) = (\beta_0(\tau), \beta_1(\tau))'$, $\beta_0(\tau) = q_0(\tau)$, and $\beta_1(\tau) = q(\tau)$. We estimate $\beta(\tau)$ by $\hat{\beta}(\tau)$, where

$$\hat{\beta}(\tau) = \underset{b=(b_0,b_1)'\in\Re^2}{\arg\min} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \dot{A}_i' b \right),$$

$\dot{A}_i = (1, A_i)'$, and $\rho_\tau(u) = u(\tau - 1\{u \le 0\})$ is the standard check function. We refer to $\hat{\beta}_1(\tau)$, the second element of $\hat{\beta}(\tau)$, as our SQR estimator for the τ-th QTE. As A_i is a dummy variable, $\hat{\beta}_1(\tau)$ is numerically the same as the difference between the τ-th empirical quantiles of Y in the treatment and control groups.

Theorem 1.3.1. *If Assumptions 1(i)–1(iii) and 2 hold, then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}\left(\hat{\beta}_1(\tau) - q(\tau)\right) \rightsquigarrow \mathcal{B}_{sqr}(\tau), \ \text{as } n \to \infty,$$

where $\mathcal{B}_{sqr}(\cdot)$ is a Gaussian process with covariance kernel $\Sigma_{sqr}(\cdot,\cdot)$. The expression for $\Sigma_{sqr}(\cdot,\cdot)$ can be found in the Appendix.

The asymptotic variance for $\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right)$ is $\zeta_Y^2(\pi,\tau) + \zeta_A^2(\pi,\tau) + \zeta_S^2(\tau)$, where

$$\zeta_Y^2(\pi,\tau) = \frac{\tau(1-\tau) - \mathbb{E}m_1^2(S,\tau)}{\pi f_1^2(q_1(\tau))} + \frac{\tau(1-\tau) - \mathbb{E}m_0^2(S,\tau)}{(1-\pi) f_0^2(q_0(\tau))},$$

$$\zeta_A^2(\pi,\tau) = \mathbb{E}\gamma(S)\left(\frac{m_1(S,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(S,\tau)}{(1-\pi) f_0(q_0(\tau))}\right)^2,$$

$$\zeta_S^2(\tau) = \mathbb{E}\left(\frac{m_1(S,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S,\tau)}{f_0(q_0(\tau))}\right)^2,$$

and $m_j(s,\tau) = \mathbb{E}(\tau - 1\{Y(j) \le q_j(\tau)\} | S = s)$. Note that, if the treatment assignment rule achieves strong balance or the stratification is irrelevant[5] then $\zeta_A^2(\pi,\tau) = 0$.

[5] It means $\mathbb{P}(Y(j) \le q_j(\tau) | S = s) = \tau$ for $s \in \mathcal{S}, j = 0, 1$.

1.3.2 Inverse Propensity Score weighted Quantile Regression

Denote $\hat{\pi}(s) = n_1(s)/n(s)$, $n_1(s) = \sum_{i=1}^n A_i 1\{S_i = s\}$, and $n(s) = \sum_{i=1}^n 1\{S_i = s\}$. Note $\hat{\pi}(S_i)$ is an estimator for the propensity score, i.e., π. In addition, Assumption 1(ii) implies that the unconfoundedness condition holds. Thus, following the lead of Firpo (2007), we can estimate $q_j(\tau)$ by the IPW QR. Let

$$\hat{q}_1(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{A_i}{\hat{\pi}(S_i)}\rho_\tau(Y_i - q) \quad \text{and} \quad \hat{q}_0(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{1-A_i}{1-\hat{\pi}(S_i)}\rho_\tau(Y_i - q).$$

We then estimate $q(\tau)$ by $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$.

Theorem 1.3.2. *If Assumptions 1(i), 1(ii), 1(iv) and 2 hold, then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}\left(\hat{q}(\tau) - q(\tau)\right) \rightsquigarrow \mathcal{B}_{ipw}(\tau), \text{ as } n \to \infty,$$

where $\mathcal{B}_{ipw}(\cdot)$ is a scalar Gaussian process with covariance kernel $\Sigma_{ipw}(\cdot,\cdot)$. The expression for $\Sigma_{ipw}(\cdot,\cdot)$ can be found in the Appendix.

Two remarks are in order. First, the asymptotic variance for $\hat{q}(\tau)$ is

$$\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau).$$

When strong balance is not achieved and the stratification is relevant, we have $\zeta_A^2(\pi,\tau) > 0$. Thus, $\hat{q}(\tau)$ is more efficient than $\hat{\beta}_1(\tau)$ in the sense that

$$\Sigma_{ipw}(\tau,\tau) < \Sigma_{sqr}(\tau,\tau).$$

When strong balance is achieved ($\gamma(s) = 0$), we have $\zeta_A^2(\pi,\tau) = 0$. Thus, the two estimators are asymptotically first-order equivalent. Based on the same argument, one can potentially prove that, when strong balance is not achieved and the stratification is relevant, the IPW estimator for ATE has strictly smaller asymptotic variance than the simple two-sample difference and strata fixed effects estimators studied by Bugni et al. (2018), and is asymptotically equivalent to the fully saturated linear regression estimator proposed by Bugni et al. (2019). Second, since the amount of "balance" of the treatment assignment rule does not play a role in the limiting distribution of the IPW estimator, Assumption 1(iii) is replaced by Assumption 1(iv).

1.4 Weighted Bootstrap

In this section, we approximate the asymptotic distributions of the SQR and IPW estimators via the weighted bootstrap. Let $\{\xi_i\}_{i=1}^n$ be a sequence of bootstrap weights which will be specified later. Further denote $n_1^w(s) = \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}$, $n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\}$, and $\hat{\pi}^w(s) = n_1^w(s)/n^w(s)$. The weighted bootstrap counterparts for the two estimators we study in this paper can then be written respectively as

$$\hat{\beta}^w(\tau) = \arg\min_b \sum_{i=1}^n \xi_i \rho_\tau \left(Y_i - \dot{A}_i' b\right)$$

and

$$\hat{q}^w(\tau) = \hat{q}_1^w(\tau) - \hat{q}_0^w(\tau),$$

where

$$\hat{q}_1^w(\tau) = \arg\min_q \sum_{i=1}^n \frac{\xi_i A_i}{\hat{\pi}^w(S_i)} \rho_\tau\left(Y_i - q\right) \quad \text{and} \quad \hat{q}_0^w(\tau) = \arg\min_q \sum_{i=1}^n \frac{\xi_i(1 - A_i)}{1 - \hat{\pi}^w(S_i)} \rho_\tau\left(Y_i - q\right).$$

The second element $\hat{\beta}_1^w(\tau)$ of $\hat{\beta}^w(\tau)$ and $\hat{q}^w(\tau)$ are the SQR and IPW bootstrap estimators for the τ-th QTE, respectively. Next, we specify the bootstrap weights.

Assumption 3. *Suppose $\{\xi_i\}_{i=1}^n$ is a sequence of nonnegative i.i.d. random variables with unit expectation and variance and a sub-exponential upper tail.*

The nonnegativity is required to maintain the convexity of the quantile regression objective function. The other conditions in Assumption 3 are common for the weighted bootstrap approximation. In practice, we generate $\{\xi_i\}_{i=1}^n$ by the standard exponential distribution. The corresponding weighted bootstrap is also known as the Bayesian bootstrap.

Theorem 1.4.1. *If Assumptions 1(i)–1(iii), 2, and 3 hold, then uniformly over $\tau \in \Upsilon$ and conditionally on data,*

$$\sqrt{n}\left(\hat{\beta}_1^w(\tau) - \hat{\beta}_1(\tau)\right) \rightsquigarrow \tilde{\mathcal{B}}_{sqr}(\tau), \text{ as } n \to \infty,$$

where $\tilde{\mathcal{B}}_{sqr}(\tau)$ is a Gaussian process. In addition, $\tilde{\mathcal{B}}_{sqr}(\tau)$ shares the same covariance kernel with $\mathcal{B}_{sqr}(\tau)$ defined in Theorems 1.3.1 with $\gamma(s)$ there replaced by $\pi(1-\pi)$.

If Assumptions 1(i), 1(ii), 1(iv), 2, 3 hold, then uniformly over $\tau \in \Upsilon$ and conditionally *on data,*

$$\sqrt{n}\left(\hat{q}^w(\tau) - \hat{q}(\tau)\right) \rightsquigarrow \mathcal{B}_{ipw}(\tau), \ \textit{as } n \to \infty,$$

where $\mathcal{B}_{ipw}(\tau)$ is the same Gaussian process defined in Theorem 1.3.2.

Four remarks are in order. First, the weighted bootstrap sample does not preserve the negative cross-sectional dependence in the original sample. Asymptotic variances of the weighted bootstrap estimators equal those of their original sample counterparts as if SRS is applied. In fact, the asymptotic variance for $\hat{\beta}_1^w(\tau)$ is

$$\zeta_Y^2(\pi,\tau) + \tilde{\zeta}_A^2(\pi,\tau) + \zeta_S^2(\tau),$$

where

$$\tilde{\zeta}_A^2(\pi,\tau) = \mathbb{E}\pi(1-\pi)\left(\frac{m_1(S,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(S,\tau)}{(1-\pi)f_0(q_0(\tau))}\right)^2.$$

This asymptotic variance is intuitive as the weight ξ_i is independent with each other, which implies that, conditionally on data, the bootstrap sample observations are independent. As $\gamma(s) \leq \pi(1-\pi)$, we have

$$\zeta_A^2(\pi,\tau) \leq \tilde{\zeta}_A^2(\pi,\tau).$$

If the inequality is strict, then the weighted bootstrap overestimates the asymptotic variance of the SQR estimator, and thus, the Wald test constructed using the SQR estimator and its weighted bootstrap standard error is conservative.

Second, the asymptotic distribution of the weighted bootstrap IPW estimator coincides with that of the original estimator. The asymptotic size of the Wald test constructed using the IPW estimator and its weighted bootstrap standard error then equals the nominal level. Theorem 1.3.2 shows that the asymptotic variance for $\hat{q}(\tau)$ is invariant in the treatment assignment rule applied. Thus, even though the weighted bootstrap sample ignores the cross-sectional dependence and behaves as if the treatment status is generated randomly, the asymptotic variance for $\hat{q}^w(\tau)$ is still

$$\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau).$$

Third, the validity of weighted bootstrap for the IPW estimator only requires Assumption 1(iv) instead of 1(iii), for the same reason mentioned after Theorem 1.3.2.

Fourth, it is possible to consider the conventional nonparametric bootstrap which generates the bootstrap sample from the empirical distribution of the data. If the observations are i.i.d., van der Vaart and Wellner (1996, Section 3.6) show that the conventional bootstrap is first-order equivalent to a weighted bootstrap with Poisson(1) weights. However, in the current setting, $\{A_i\}_{i\geq 1}$ is dependent. It is technically challenging to rigorously show that the above equivalence still holds. We leave it as an interesting topic for future research.

1.5 Covariate-Adaptive Bootstrap

In this section, we consider the covariate-adaptive bootstrap procedure as follows:

(i) Draw $\{S_i^*\}_{i=1}^n$ from the empirical distribution of $\{S_i\}_{i=1}^n$ with replacement.

(ii) Generate $\{A_i^*\}_{i=1}^n$ based on $\{S_i^*\}_{i=1}^n$ and the treatment assignment rule.

(iii) For $A_i^* = a$ and $S_i^* = s$, draw Y_i^* from the empirical distribution of Y_i given $A_i = a$ and $S_i = s$ with replacement.

First, Step (i) is the conventional nonparametric bootstrap. The bootstrap sample $\{S_i^*\}_{i=1}^n$ is obtained by drawing from the empirical distribution of $\{S_i\}_{i=1}^n$ with replacement n times. Second, Step (ii) follows the treatment assignment rule, and thus preserves the cross-sectional dependence structure in the bootstrap sample, even after conditioning on data. The weighted bootstrap sample, by contrast, is cross-sectionally independent given data. Third, Step (iii) applies the conventional bootstrap procedure to the outcome Y_i in the cell $(S_i, A_i) = (s, a) \in \mathcal{S} \times \{0, 1\}$. Given that the original data contain $n_a(s)$ observations in this cell, in this step, the bootstrap sample $\{Y_i^*\}_{i:A_i^*=a, S_i^*=s}$ is obtained by drawing from the empirical distribution of these $n_a(s)$ outcomes with replacement $n_a^*(s)$ times, where $n_a^*(s) = \sum_{i=1}^n 1\{A_i^* = a, S_i^* = s\}$. Unlike the conventional bootstrap, here both $n_a(s)$ and $n_a^*(s)$ are random and are not necessarily the same. Last, to implement the covariate-adaptive bootstrap, researchers need to know the treatment assignment rule for the original sample. Unlike in observational studies, such information is usually available for RCTs. If one only knows that the treatment assignment rule achieves strong balance, then Theorem 1.5.1 below still holds, provided that the bootstrap sample is generated from any treatment assignment rule that achieves strong balance. Even worse, if no information on the treatment assignment rule is available, then one cannot implement the covariate-adaptive bootstrap inference. In this case, the weighted bootstrap for the IPW estimator can still provide a non-conservative Wald test, as shown in Theorem 1.4.1.

Using the bootstrap sample $\{Y_i^*, A_i^*, S_i^*\}_{i=1}^n$, we can estimate QTE by the two methods considered in the paper. Let $n_1^*(s) = \sum_{i=1}^n A_i^* 1\{S_i^* = s\}$, $n^*(s) = \sum_{i=1}^n 1\{S_i^* = s\}$, $\hat{\pi}^*(s) = \frac{n_1^*(s)}{n^*(s)}$, and $\dot{A}_i^* = (1, A_i^*)'$. Then, the two bootstrap estimators can be written respectively as

$$\hat{\beta}^*(\tau) = \arg\min_b \sum_{i=1}^n \rho_\tau \left(Y_i^* - \dot{A}_i^* b\right)$$

and

$$\hat{q}^*(\tau) = \hat{q}_1^*(\tau) - \hat{q}_0^*(\tau),$$

where

$$\hat{q}_1^* = \arg\min_q \sum_{i=1}^n \frac{A_i^*}{\hat{\pi}^*(S_i^*)} \rho_\tau(Y_i^* - q) \quad \text{and} \quad \hat{q}_0^* = \arg\min_q \sum_{i=1}^n \frac{1 - A_i^*}{1 - \hat{\pi}^*(S_i^*)} \rho_\tau(Y_i^* - q).$$

The second element $\hat{\beta}_1^*(\tau)$ of $\hat{\beta}^*(\tau)$ and $\hat{q}^*(\tau)$ are the SQR and IPW bootstrap estimators for the τ-th QTE, respectively. Parallel to Assumption 1, we make the following assumption for the bootstrap sample.

Assumption 4. *Let* $D_n^*(s) = \sum_{i=1}^n (A_i^* - \pi) 1\{S_i^* = s\}$.

(i) $\left\{ \left\{ \frac{D_n^*(s)}{\sqrt{n}} \right\}_{s \in \mathcal{S}} \middle| \{S_i^*\}_{i=1}^n \right\} \rightsquigarrow N(0, \Sigma_D)$ *a.s., where* $\Sigma_D = diag\{p(s)\gamma(s) : s \in \mathcal{S}\}$.

(ii) $\sup_{s \in \mathcal{S}} \frac{|D_n^*(s)|}{\sqrt{n^*(s)}} = O_p(1)$, $\sup_{s \in \mathcal{S}} \frac{|D_n(s)|}{\sqrt{n(s)}} = O_p(1)$.

Assumption 4(i) is a high-level assumption. Obviously, it holds for SRS. For WEI, this condition holds by the same argument in Bugni et al. (2018, Lemma B.12) with the fact that $\frac{n^*(s)}{n(s)} \xrightarrow{p} 1$. For BCD, as shown in Bugni et al. (2018, Lemma B.11),

$$D_n^*(s) | \{S_i^*\}_{i=1}^n = O_p(1).$$

Therefore, $D_n^*(s)/\sqrt{n^*(s)} \xrightarrow{p} 0$ and Assumption 4(i) holds with $\gamma(s) = 0$. For SBR, it is clear that $|D_n^*(s)| \le 1$. Thus, Assumption 4(i) holds with $\gamma(s) = 0$ as well. In addition, as $p(s) > 0$, based on the standard bootstrap results, we have $n^*(s)/n \xrightarrow{p} p(s)$ and $n(s)/n \xrightarrow{p} p(s)$. Therefore, Assumption 4(i) is sufficient for Assumption 4(ii). Last, note that Assumption 4(ii) implies Assumption 1(iv).

Theorem 1.5.1. *Suppose Assumptions 1(i), 1(ii), 2, and 4(ii) hold. Then, uniformly over* $\tau \in \Upsilon$ *and conditionally on data,*

$$\sqrt{n}(\hat{q}^*(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}_{ipw}(\tau), \ \ as \ n \to \infty.$$

If, in addition, Assumptions 1(iii) and 4(i) hold, then

$$\sqrt{n}\left(\hat{\beta}_1^*(\tau) - \hat{q}(\tau)\right) \rightsquigarrow \mathcal{B}_{sqr}(\tau), \ \ as \ n \to \infty.$$

Here, $\mathcal{B}_{sqr}(\tau)$ *and* $\mathcal{B}_{ipw}(\tau)$ *are two Gaussian processes defined in Theorem 1.3.1 and 1.3.2, respectively.*

Several remarks are in order. First, unlike the usual bootstrap estimator, the covariate-adaptive bootstrap SQR estimator is not centered around its corresponding counterpart from the original sample, but rather $\hat{q}(\tau)$. The reason is that the treatment status A_i^* is not generated by bootstrap. In the linear expansion for the bootstrap estimator $\hat{\beta}_1^*(\tau)$, the part of the influence function that accounts for the variation generated by A_i^* need not be centered. We also know from the proof of Theorem 1.3.2 that $\hat{q}(\tau)$ do not have an influence function that represents the variation generated by A_i. Thus, $\hat{q}(\tau)$ can be used to center $\hat{\beta}_1^*(\tau)$.

Second, the choice of $\hat{q}(\tau)$ as the center is somehow ad-hoc. In fact, any estimator $\tilde{q}(\tau)$ that is first-order equivalent to $\hat{q}(\tau)$ in the sense that

$$\sup_{\tau \in \Upsilon} |\tilde{q}(\tau) - \hat{q}(\tau)| = o_p(1/\sqrt{n})$$

can serve as the center for the bootstrap estimators $\hat{q}^*(\tau)$ and $\hat{\beta}_1^*(\tau)$.

Third, when the treatment assignment rule achieves strong balance, $\hat{\beta}_1(\tau)$ and $\hat{q}(\tau)$ are first-order equivalent. In this case, $\hat{\beta}_1(\tau)$ can serve as the center for $\hat{\beta}_1^*(\tau)$ and various bootstrap inference methods are valid. On the other hand, when the treatment assignment rule does not achieve strong balance, $\hat{\beta}_1(\tau)$ and $\hat{q}(\tau)$ are not first-order equivalent. In this case, the asymptotic size of the percentile bootstrap for the SQR estimator using the quantiles of $\hat{\beta}_1^*(\tau)$ does not equal the nominal level. In the next section, we propose a way to compute the bootstrap standard error which does not depend on the choice of the center. Based on the bootstrap standard error, researchers can construct t-statistics and use standard normal critical values for inference.

Fourth, for ATE, we can use the same bootstrap sample to compute the standard errors for the simple and strata fixed effects estimators proposed in Bugni et al. (2018) as well as the IPW estimator. We expect that all the results in this paper hold for the ATE as well.

1.6 Simulation

We can summarize four bootstrap scenarios from the analysis in Sections 1.4 and 1.5: (i) the SQR estimator with the weighted bootstrap, (ii) the IPW estimator with either the weighted or covariate-adaptive bootstrap, (iii) the SQR estimator with the covariate-adaptive bootstrap when the assignment rule achieves strong balance, and (iv) the SQR estimator with the covariate-adaptive bootstrap when the assignment rule does not achieve strong balance. The results of Sections 1.4 and 1.5 imply that the bootstrap in scenario (i) produces conservative Wald-tests when the treatment assignment rule is not SRS. For scenarios (ii) and (iii), various bootstrap based inference methods are valid. However, for scenario (iv), researchers should be careful about the centering issue. In particular, the percentile bootstrap inference using the quantiles of $\hat{\beta}_1^*$ is invalid. In the following, we propose one single bootstrap inference method that works for scenarios (ii)–(iv). In addition, the proposed method does not require the knowledge of the centering.

We take the IPW estimator as an example. We can repeat the bootstrap estimation[6] B times and obtain B bootstrap IPW estimates, denoted as $\{\hat{q}_b^*(\tau)\}_{b=1}^B$. Further denote $\hat{Q}(\alpha)$ as the α-th empirical quantile of $\{\hat{q}_b^*(\tau)\}_{b=1}^B$. We can test the null hypothesis that $q(\tau) = q^0(\tau)$ via $1\left\{\left|\frac{\hat{q}(\tau)-q^0(\tau)}{\hat{\sigma}_n^*}\right| > z_{1-\alpha/2}\right\}$, where $\hat{q}(\tau)$, $z_{1-\alpha/2}$, and $\hat{\sigma}_n^*$ are the IPW estimator, the $(1-\alpha/2)$-th quantile of the standard normal distribution, and

$$\hat{\sigma}_n^* = \frac{\hat{Q}(0.975) - \hat{Q}(0.025)}{z_{0.975} - z_{0.025}},$$

respectively. In scenarios (ii)–(iv), the asymptotic size of such test equals the nominal level α. In scenarios (ii) and (iii), we recommend the t-statistic and confidence interval using this particular bootstrap standard error (i.e., $\hat{\sigma}_n^*$) over other bootstrap inference methods (e.g., bootstrap confidence interval, percentile bootstrap confidence interval, etc.) because based on unreported simulations, they have better finite sample performance.

1.6.1 Data Generating Processes

We consider two DGPs with parameters $\gamma = 4$, $\sigma = 2$, and μ which will be specified later.

(i) Let Z be standardized Beta$(2,2)$ distributed, $S_i = \sum_{j=1}^4 1\{Z_i \le g_j\}$, and $(g_1, \cdots, g_4) =$

[6] For the IPW estimator, we can use either the weighted or covariate-adaptive bootstrap. For the SQR estimator, we can only use the covariate-adaptive bootstrap.

$(-0.25\sqrt{20}, 0, 0.25\sqrt{20}, 0.5\sqrt{20})$. The outcome equation is

$$Y_i = A_i\mu + \gamma Z_i + \eta_i,$$

where $\eta_i = \sigma A_i \varepsilon_{i,1} + (1 - A_i)\varepsilon_{i,2}$ and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

(ii) Let Z be uniformly distributed on $[-2, 2]$, $S_i = \sum_{j=1}^{4} 1\{Z_i \le g_j\}$, and $(g_1, \cdots, g_4) = (-1, 0, 1, 2)$. The outcome equation is

$$Y_i = A_i\mu + A_i\nu_{i,1} + (1 - A_i)\nu_{i,0} + \eta_i,$$

where $\nu_{i,0} = \gamma Z_i^2 1\{|Z_i| \ge 1\} + \frac{\gamma}{4}(2 - Z_i^2)1\{|Z_i| < 1\}$, $\nu_{i,1} = -\nu_{i,0}$, $\eta_i = \sigma(1 + Z_i^2)A_i\varepsilon_{i,1} + (1 + Z_i^2)(1 - A_i)\varepsilon_{i,2}$, and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are mutually independent $T(3)/3$ distributed.

When $\pi = \frac{1}{2}$, for each DGP, we consider four randomization schemes:

(i) SRS: Treatment assignment is generated as in Example 1.

(ii) WEI: Treatment assignment is generated as in Example 2 with $\phi(x) = (1 - x)/2$.

(iii) BCD: Treatment assignment is generated as in Example 3 with $\lambda = 0.75$.

(iv) SBR: Treatment assignment is generated as in Example 4.

When $\pi \ne 0.5$, BCD is not defined while WEI is not defined in the original paper (Wei, 1978). Recently, Hu (2016) generalizes the adaptive biased-coin design (i.e., WEI) to multiple treatment values and unequal target treatment ratios. Here, for $\pi \ne 0.5$, we only consider SRS and SBR as in Bugni et al. (2018). We conduct the simulations with sample sizes $n = 200$ and 400. The numbers of simulation replications and bootstrap samples are 1000. Under the null, $\mu = 0$ and we compute the true parameters of interest using simulations with 10^6 sample size and 10^4 replications. Under the alternative, we perturb the true values by $\mu = 1$ and $\mu = 0.75$ for $n = 200$ and 400, respectively. We report the results for the median QTE. Section 1.11.6 contains additional simulation results for ATE and QTEs with $\tau = 0.25$ and 0.75. All the observations made in this section still apply.

1.6.2 QTE, $\pi = 0.5$

We consider the Wald test with six t-statistics and 95% nominal rate. We construct the t-statistics using one of our two point estimates and some estimate of the standard error. We will reject the null hypothesis when the absolute value of the t-statistic is greater than 1.96. The details about the point estimates and standard errors are as follows:

(i) "s/naive": the point estimator is computed by the SQR and its standard error $\hat{\sigma}_{naive}(\tau)$ is computed as

$$\begin{aligned}\hat{\sigma}^2_{naive}(\tau) =& \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_1^2(S_i,\tau)}{\pi \hat{f}_1^2(\hat{q}_1(\tau))} + \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_0^2(S_i,\tau)}{(1-\pi)\hat{f}_0^2(\hat{q}_0(\tau))} \\ &+ \frac{1}{n}\sum_{i=1}^n \pi(1-\pi)\left(\frac{\hat{m}_1(S_i,\tau)}{\pi \hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{m}_0(S_i,\tau)}{(1-\pi)\hat{f}_0(\hat{q}_0(\tau))}\right)^2 \\ &+ \frac{1}{n}\sum_{i=1}^n \left(\frac{\hat{m}_1(S_i,\tau)}{\hat{f}_1(\hat{q}_1(\tau))} - \frac{\hat{m}_0(S_i,\tau)}{\hat{f}_0(\hat{q}_0(\tau))}\right)^2, \end{aligned} \tag{1.6.1}$$

where $\hat{q}_j(\tau)$ is the τ-th empirical quantile of $Y_i|A_i = j$,

$$\hat{m}_{i,1}(s,\tau) = \frac{\sum_{i=1}^n A_i 1\{S_i = s\}(\tau - 1\{Y_i \leq \hat{q}_1(\tau)\})}{n_1(s)},$$

$$\hat{m}_{i,0}(s,\tau) = \frac{\sum_{i=1}^n (1-A_i) 1\{S_i = s\}(\tau - 1\{Y_i \leq \hat{q}_0(\tau)\})}{n(s) - n_1(s)}.$$

For $j = 0, 1$, $\hat{f}_j(\cdot)$ is computed by the kernel density estimation using the observations Y_i provided that $A_i = j$, bandwidth $h_j = 1.06\hat{\sigma}_j n_j^{-1/5}$, Gaussian kernel function, standard deviation $\hat{\sigma}_j$ of the observations Y_i provided that $A_i = j$, and $n_j = \sum_{i=1}^n 1\{A_i = j\}$.

(ii) "s/adj": exactly the same as the "s/naive" method with one difference: replacing $\pi(1-\pi)$ in (1.6.1) by $\gamma(S_i)$.

(iii) "s/W": the point estimator is computed by the SQR and its standard error $\hat{\sigma}_W(\tau)$ is computed by the weighted bootstrap procedure. The bootstrap weights $\{\xi_i\}_{i=1}^n$ are generated from the standard exponential distribution. Denote $\{\hat{\beta}^w_{1,b}\}_{b=1}^B$ as the collection of B weighted bootstrap SQR estimates. Then,

$$\hat{\sigma}_W(\tau) = \frac{\hat{Q}(0.975) - \hat{Q}(0.025)}{z_{0.975} - z_{0.025}},$$

where $\hat{Q}(\alpha)$ is the α-th empirical quantile of $\{\hat{\beta}^w_{1,b}(\tau)\}_{b=1}^B$.

(iv) "ipw/W": the same as above with one difference: the estimation method for both the original and bootstrap samples is the IPW QR.

(v) "s/CA": the point estimator is computed by the SQR and its standard error $\hat{\sigma}_{CA}(\tau)$ is computed by the covariate-adaptive bootstrap procedure. Denote $\{\hat{\beta}^*_{1,b}\}_{b=1}^B$ as the collection of B estimates obtained by the SQR applied to the samples generated by the covariate-adaptive bootstrap procedure. Then,

$$\hat{\sigma}_{CA}(\tau) = \frac{\hat{Q}(0.975) - \hat{Q}(0.025)}{z_{0.975} - z_{0.025}},$$

where $\hat{Q}(\alpha)$ is the α-th empirical quantile of $\{\hat{\beta}^*_{1,b}(\tau)\}_{b=1}^B$.

(vi) "ipw/CA": the same as above with one difference: the estimation method for both the original and bootstrap samples is the IPW QR.

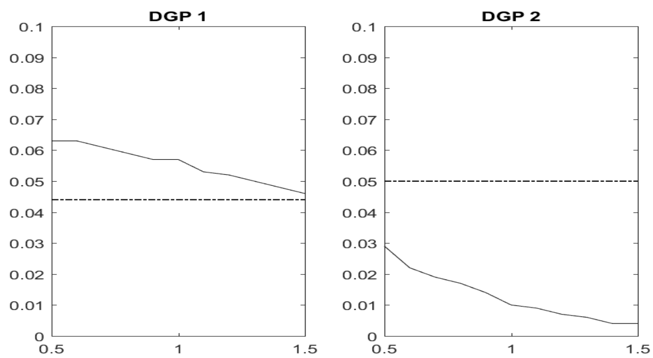
Tables I and II present the rejection probabilities (multiplied by 100) for the six t-tests under both the null hypothesis and the alternative hypothesis, with sample sizes $n = 200$ and 400, respectively. In these two tables, columns M and A represent DGPs and treatment assignment rules, respectively. From the rejection probabilities under the null, we can make five observations. First, the naive t-test ("s/naive") is conservative for WEI, BCD, and SBR, which is consistent with the findings for ATE estimators by Shao et al. (2010) and Bugni et al. (2018). Second, although the asymptotic size of the adjusted t-test ("s/adj") is expected to equal the nominal level, it does not perform well for DGP2. The main reason is that, in order to analytically compute the standard error, we must compute nuisance parameters such as the unconditional densities of $Y(0)$ and $Y(1)$, which requires tuning parameters. We further compute the standard errors following (1.6.1) with $\pi(1-\pi)$ and the tuning parameter h_j replaced by $\gamma(S_i)$ and $1.06 C_f \hat{\sigma}_j n_j^{-1/5}$, respectively, for some constant $C_f \in [0.5, 1.5]$. Figure 1.1 plots the rejection probabilities of the "s/adj" t-tests against C_f for the BCD assignment rule with $n = 200$, $\tau = 0.5$, and $\pi = 0.5$. We see that (i) the rejection probability is sensitive to the choice of bandwidth, (ii) there is no universal optimal bandwidth across two DGPs, and (iii) the covariate-adaptive bootstrap t-tests ("s/CA") represented by the dotted dash lines are quite stable across different DGPs and close to the nominal rate of rejection. Third, the weighted bootstrap t-test for the SQR estimator ("s/W") is conservative, especially for the BCD and SBR assignment rules which achieve strong balance. Fourth, the rejection probabilities of the weighted bootstrap t-test for the IPW estimator ("ipw/W") are close to the nominal rate even for sample size $n = 200$, which is consistent with Theorem 1.4.1. Last, the rejection rates for the two covariate-adaptive bootstrap t-tests ("s/CA" and "ipw/CA") are close to the nominal rate, which is consistent with Theorem 1.5.1.

Table I. $n = 200, \tau = 0.5, \pi = 0.5$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	4.5	4.5	4.7	4.4	4.4	3.9	18.3	18.3	19.3	44.1	20.0	42.9
	WEI	1.2	4.0	1.4	4.3	3.7	3.5	11.6	29.5	13.8	44.7	29.8	43.6
	BCD	0.2	5.7	0.3	4.1	4.4	3.9	7.2	47.2	9.5	45.3	43.4	44.8
	SBR	0.1	5.7	0.1	4.6	4.5	4.4	8.5	48.5	9.9	46.0	45.7	44.8
2	SRS	0.4	0.4	4.7	5.2	5.2	5.3	79.7	79.7	90.4	91.6	90.2	91.3
	WEI	0.6	0.6	4.5	5.8	5.2	5.7	80.2	80.7	90.7	90.9	91.3	90.6
	BCD	1.0	1.0	4.5	5.1	5.0	5.3	79.6	80.4	90.2	91.1	90.8	90.6
	SBR	0.8	1.1	4.8	5.3	4.6	4.7	77.1	77.4	89.7	90.1	89.9	89.9

Table II. $n = 400, \tau = 0.5, \pi = 0.5$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	4.2	4.2	5.4	4.0	4.6	4.1	21.8	21.8	23.2	50.2	23.5	50.2
	WEI	1.0	4.9	0.8	4.7	4.6	4.2	14.7	35.6	16.0	50.3	35.0	50.7
	BCD	0.3	4.5	0.2	4.3	3.5	4.0	8.9	52.6	11.7	50.2	49.3	49.6
	SBR	0.2	4.6	0.0	3.7	3.6	3.7	8.9	55.0	10.9	51.8	52.4	51.9
2	SRS	1.2	1.2	4.3	4.8	4.6	5.0	89.7	89.7	95.6	95.6	95.7	95.7
	WEI	1.4	1.6	5.7	6.0	5.5	5.7	89.2	89.2	95.4	94.8	95.1	94.8
	BCD	1.3	1.3	5.5	6.1	5.1	5.2	88.7	88.9	95.2	95.4	95.7	95.6
	SBR	0.6	0.6	4.0	3.9	3.8	3.8	90.0	90.2	95.4	95.4	95.8	95.7



Note: Rejection probabilities for BCD assignment rule with $n = 200$, $\pi = 0.5$, and $\tau = 0.5$. The X-axis is C_f. The solid lines are the rejection probabilities for "s/adj". The densities of Y_j is computed using the tuning parameters $h_j = 1.06 C_f \hat{\sigma}_j n_j^{-1/5}$, for $j = 0, 1$. The dotted dash lines are the rejection probability for "s/CA".

Figure 1.1. Rejection Probabilities Across Different Bandwidth Values

Turning to the rejection rates under the alternative in Tables I and II, we can make two additional observations. First, for BCD and SBR, the rejection probabilities (power) for "ipw/W", "s/CA", and "ipw/CA" are close. This is because both BCD and SBR achieve strong balance. In this case, the two estimators we propose are asymptotically first-order equivalent. Second, for DGP1 with SRS and WEI assignment rules, "ipw/CA" is more powerful than "s/CA". This confirms our theoretical finding that the IPW estimator is *strictly* more efficient than the SQR estimator when the treatment assignment rule does *not* achieve strong balance. For DGP2 the three t-tests, i.e., "ipw/W", "s/CA", and "ipw/CA", have similar power.

1.6.3 QTE, $\pi = 0.7$

Tables III and IV show the similar results with $\pi = 0.7$. The same comments for Tables I and II still apply.

Table III. $n = 200, \tau = 0.5, \pi = 0.7$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	4.8	4.8	5.2	4.7	3.4	4.4	17.0	17.0	17.2	42.5	16.7	40.6
	SBR	0.1	0.7	0.2	4.0	4.4	3.7	4.3	21.2	6.0	45.5	45.7	43.4
2	SRS	1.6	1.6	5.2	5.4	5.1	5.3	77.1	77.1	89.1	90.3	89.5	89.4
	SBR	0.4	0.5	3.9	4.8	4.5	4.8	76.0	76.9	89.2	91.1	90.1	90.0

Table IV. $n = 400, \tau = 0.5, \pi = 0.7$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	4.4	4.4	5.1	3.9	4.8	3.7	18.4	18.4	18.7	47.9	19.4	46.6
	SBR	0.1	0.2	0	3.9	3.5	4	4.2	22	5.9	49.8	50.5	48.2
2	SRS	0.7	0.7	3.9	4.2	4.2	4.7	86.7	86.7	93.9	93.3	94.1	93.6
	SBR	0.6	0.6	3.5	3.6	3.7	3.7	88.3	88.8	94.8	95.2	95.5	95.2

1.6.4 Difference between Two QTEs

Last, we consider to infer $q(0.25) - q(0.75)$ when $\pi = 0.5$:

$$H_0: \ q(0.25) - q(0.75) = \text{the true value} \quad \text{v.s.} \quad H_1: \ q(0.25) - q(0.75) = \text{the true value} + \mu,$$

where $\mu = 1$ and 0.75 for sample sizes 200 and 400, respectively. The two estimators for QTEs at $\tau = 0.25$ and 0.75 are correlated. We can compute the naive and adjusted standard errors for the SQR estimator by taking this covariance structure into account.[7] On the other hand, in addition to avoiding the tuning parameters, another advantage of the bootstrap inference is

[7] The formulas for the covariances can be found in the proofs of Theorems 1.3.1 and 1.3.2.

it does not require the knowledge of this complicated covariance structure. Researchers may construct the t-statistic using the difference of two QTE estimators with the corresponding weighted and covariate-adaptive bootstrap standard errors, which are calculated using the exact same procedure as in Sections 1.4 and 1.5. Taking the SQR estimator as an example, we estimate $q(0.25)-q(0.75)$ via $\hat{\beta}_1(0.25)-\hat{\beta}_1(0.75)$ and the corresponding covariate-adaptive bootstrap standard error is

$$\hat{\sigma}_{CA} = \frac{\hat{Q}(0.975) - \hat{Q}(0.025)}{z_{0.975} - z_{0.025}},$$

where $\hat{Q}(\alpha)$ is the α-th empirical quantile of $\{\hat{\beta}^*_{1,b}(0.25) - \hat{\beta}^*_{1,b}(0.75)\}_{b=1}^B$.

Based on the rejection rates reported in Tables V and VI, the general observations for the previous simulation results still apply. Although under the null, the rejection rates for "ipw/W", "S/CA", "ipw/CA" in DGP2 are below the nominal 5%, they gradually increase as the sample size increases from 200 to 400.

Table V. $n = 200, q(0.25) - q(0.75)$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	4.0	4.0	3.6	3.8	3.5	3.5	15.6	15.6	14.9	19.4	16.0	19.4
	WEI	2.3	4.9	2.0	4.0	5.1	3.9	11.3	17.9	11.0	19.0	16.0	18.6
	BCD	1.0	4.1	1.1	4.4	3.7	4.2	9.9	20.7	10.1	22.0	20.6	21.4
	SBR	1.1	4.3	0.9	4.1	4.1	4.2	9.4	21.8	8.7	17.3	20.0	17.2
2	SRS	5.0	5.0	3.1	3.1	3.1	3.1	53.7	53.7	47.1	48.4	47.8	48.2
	WEI	3.6	3.6	2.1	2.8	2.9	2.9	57.0	57.7	47.6	49.8	50.3	50.0
	BCD	4.2	4.8	2.4	2.5	3.6	2.7	58.0	59.4	49.1	52.0	52.8	50.8
	SBR	5.1	5.3	2.4	3.4	4.1	3.4	55.5	57.0	46.5	46.5	50.5	45.6

Table VI. $n = 400, q(0.25) - q(0.75)$

		H_0						H_1					
M	A	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA	s/naive	s/adj	s/W	ipw/W	s/CA	ipw/CA
1	SRS	3.8	3.8	3.9	5.1	3.7	5.0	17.2	17.2	15.9	21.5	16.8	21.2
	WEI	2.0	4.2	2.4	3.3	4.4	3.5	11.8	20.2	11.5	21.4	20.2	20.7
	BCD	1.4	4.4	1.4	4.3	4.4	4.1	10.5	21.8	10.2	20.7	21.5	20.6
	SBR	0.8	3.8	0.8	3.9	3.7	3.8	12.1	25.0	12.6	21.8	23.7	22.3
2	SRS	5.3	5.3	3.9	4.7	4.3	4.8	63.2	63.2	55.7	57.7	56.8	57.6
	WEI	5.4	5.8	3.4	3.7	4.1	3.5	63.6	64.4	55.6	58.0	58.0	58.5
	BCD	4.0	4.3	2.6	2.8	3.1	3.1	62.1	63.3	54.7	55.7	57.4	56.0
	SBR	5.1	5.7	4.0	4.5	4.4	4.5	61.1	62.0	52.4	51.3	56.0	53.0

1.7 Guidance for Practitioners

We recommend employing the t-statistic (or equivalently, the confidence interval) constructed using the IPW estimator and its weighted bootstrap standard error for inference in covariate-adaptive randomization, for the following four reasons. First, its asymptotic size equals the nominal level. Second, the IPW estimator has a smaller asymptotic variance than the SQR estimator when the treatment assignment rule does not achieve strong balance and the stratification is relevant.[8] Third, compared with the covariate-adaptive bootstrap, the validity of the weighted bootstrap requires a weaker condition that $\sup_{s\in\mathcal{S}}|D_n(s)/n(s)| = o_p(1)$. Fourth, this method does not require the knowledge of the exact treatment assignment rule, thus is suitable in settings where such information is lacking, e.g., using someone else's RCT or subsample analysis. When the treatment assignment rule achieves strong balance, SQR estimator can also be used. But in this case, only the covariate-adaptive bootstrap standard error is valid. Last, the Wald test using SQR estimator and the weighted bootstrap standard error is not recommended, as it is conservative when the treatment assignment rule introduces negative dependence (i.e., $\gamma(s) < \pi(1-\pi)$) such as WEI, BCD, and SBR.

1.8 Empirical Application

We illustrate our methods by estimating and inferring the average and quantile treatment effects of iron efficiency on educational attainment. The dataset we use is the same as the one analyzed by Chong et al. (2016) and Bugni et al. (2018).

1.8.1 Data Description

The dataset consists of 215 students from one Peruvian secondary school during the 2009 school year. About two thirds of students were assigned to the treatment group ($A = 1$ or $A = 2$). The other one third of students were assigned to the control group ($A = 0$). One half of the students in the treatment group were shown a video in which a physician encouraged iron supplements ($A = 1$) and the other half were shown the same encouragement from a popular soccer player ($A = 2$). Those assignments were stratified by the number of years of secondary school completed ($\mathcal{S} = \{1, \cdots, 5\}$). The field experiment used a stratified block randomization scheme with fractions $(1/3, 1/3, 1/3)$ for each group, which achieves strong balance ($\gamma(s) = 0$).

[8] In this case, for ATE, the IPW estimator also has a strictly smaller asymptotic variance than the strata fixed effects estimator studied in Bugni et al. (2018).

In the following, we focus on the observations with $A = 0$ and $A = 1$, and estimate the treatment effect of the exposure to a video of encouraging iron supplements by a physician only. This practice was also implemented in Bugni et al. (2018). In this case, the target proportions of treatment is $\pi = 1/2$. As in Chong et al. (2016), it is also possible to combine the two treatment groups, i.e., $A = 1$ and $A = 2$ and compute the treatment effects of exposure to a video of encouraging iron supplements by either a physician or a popular soccer player. Last, one can use the method developed in Bugni et al. (2019) to estimate the ATEs under multiple treatment status. However, in this setting, the estimation of QTE and the validity of bootstrap inference have not been investigated yet and are interesting topics for future research.

For each observation, we have three outcome variables: number of pills taken, grade point average, and cognitive ability measured by the average score across different Nintendo Wii games. For more details about the outcome variables, we refer interested readers to Chong et al. (2016). In the following, we focus on the grade point average only as the other two outcomes are discrete.

1.8.2 Computation

We consider three pairs of point estimates and their corresponding non-conservative standard errors: (i) the SQR estimator with the covariate-adaptive bootstrap standard error, (ii) the IPW estimator with the covariate-adaptive bootstrap standard error, and (iii) the IPW estimator with the weighted bootstrap standard error. We denote them as "s/CA", "ipw/CA", and "ipw/W", respectively. For comparison, we also compute the SQR estimator with its weighted bootstrap standard error, which is denoted as "s/W" . The SQR estimator for the τ-th QTE refers to $\hat{\beta}_1(\tau)$ as the second element of $\hat{\beta}(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}_1(\tau))$, where

$$\hat{\beta}(\tau) = \operatorname*{arg\,min}_{b=(b_0,b_1)' \in \Re^2} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \dot{A}_i' b \right),$$

$\dot{A}_i = (1, A_i)'$, and $\rho_\tau(u) = u(\tau - 1\{u \leq 0\})$ is the standard check function. It is also just the difference between the τ-th empirical quantiles of treatment and control groups. The IPW estimator refers to $\hat{q}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau)$, where

$$\hat{q}_1(\tau) = \operatorname*{arg\,min}_{q} \frac{1}{n} \sum_{i=1}^{n} \frac{A_i}{\hat{\pi}(S_i)} \rho_\tau(Y_i - q), \quad \hat{q}_0(\tau) = \operatorname*{arg\,min}_{q} \frac{1}{n} \sum_{i=1}^{n} \frac{1 - A_i}{1 - \hat{\pi}(S_i)} \rho_\tau(Y_i - q),$$

$\hat{\pi}(\cdot)$ denotes the propensity score estimator, $\hat{\pi}(s) = n_1(s)/n(s)$, $n_1(s) = \sum_{i=1}^n A_i 1\{S_i = s\}$, and $n(s) = \sum_{i=1}^n 1\{S_i = s\}$. The covariate-adaptive bootstrap standard error ("CA") refers to the standard error computed in Section 1.5. In particular, we can draw the covariate-adaptive bootstrap sample $(Y_i^*, A_i^*, S_i^*)_{i=1}^n$ following the procedure in Section 1.5. We then recompute the SQR and IPW estimates using the bootstrap sample. We repeat the bootstrap estimation B times, and obtain $\{\hat{\beta}_{b,1}^*(\tau), \hat{q}_b^*(\tau)\}_{b=1}^B$. The standard errors for SQR and IPW estimates are computed as

$$\hat{\sigma}_{sqr}(\tau) = \frac{\hat{Q}_{sqr}(0.975) - \hat{Q}_{sqr}(0.025)}{z_{0.975} - z_{0.025}} \quad \text{and} \quad \hat{\sigma}_{ipw}(\tau) = \frac{\hat{Q}_{ipw}(0.975) - \hat{Q}_{ipw}(0.025)}{z_{0.975} - z_{0.025}},$$

respectively, where $\hat{Q}_{sqr}(\alpha)$ and $\hat{Q}_{ipw}(\alpha)$ are the α-th empirical quantiles of $\{\hat{\beta}_{b,1}^*(\tau)\}_{b=1}^B$ and $\{\hat{q}_b^*(\tau)\}_{b=1}^B$, respectively, and z_α is the α-th percentile of the standard normal distribution, i.e., $z_{0.975} \approx 1.96$ and $z_{0.025} \approx -1.96$. The weighted bootstrap standard error for the IPW estimate can be computed in the same manner with only one difference, the covariate-adaptive bootstrap estimator $\{\hat{q}_b^*(\tau)\}_{b=1}^B$ is replaced by the weighted bootstrap estimator $\{\hat{q}_b^w(\tau)\}_{b=1}^B$, where for the b-th replication, $\hat{q}_b^w(\tau) = \hat{q}_{b,1}^w(\tau) - \hat{q}_{b,0}^w(\tau)$,

$$\hat{q}_{b,1}^w(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^b A_i}{\hat{\pi}^w(S_i)} \rho_\tau(Y_i - q), \hat{q}_{b,1}^w(\tau) = \arg\min_q \frac{1}{n}\sum_{i=1}^n \frac{\xi_i^b (1 - A_i)}{1 - \hat{\pi}^w(S_i)} \rho_\tau(Y_i - q),$$

$\{\xi_i^b\}_{i=1}^n$ is a sequence of i.i.d. standard exponentially distributed random variables, $\hat{\pi}^w(s) = n_1^w(s)/n^w(s)$, $n_1^w(s) = \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}$, and $n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\}$. Similarly, we compute the weighted bootstrap SQR estimates $\{\beta_{b,1}^w(\tau)\}_{b=1}^B$ as the second element of $\{\beta_b^w(\tau)\}_{b=1}^B$, where

$$\beta_b^w(\tau) = \arg\min_{b=(b_0,b_1)' \in \Re^2} \frac{1}{n}\sum_{i=1}^n \xi_i^b \rho_\tau(Y_i - \dot{A}_i' b).$$

For the ATEs, we also compute the SQR estimator with the adjusted standard error based on the analytical formula derived by Bugni et al. (2018), i.e., "s/adj". For QTE estimates, we consider quantile indexes $\{0.1, 0.15, \cdots, 0.90\}$. The number of replications for the two bootstrap methods is $B = 1000$.

1.8.3 Main Results

Table VII shows the estimates with the corresponding standard errors in parentheses. From the table, we can make several remarks. First, for both ATE and QTE, the SQR and IPW estimates are very close to each other and so do their standard errors computed via the analytical formula, weighted bootstrap, and covariate-adaptive bootstrap. This is consistent with our theory that, under strong balance, the two estimators are first-order equivalent. Second, although in theory, the weighted bootstrap standard errors for the SQR estimators should be larger than those computed via the covariate-adaptive bootstrap, in this application, they are very close. This is consistent with the finding in Bugni et al. (2018) that their adjusted p-value for the ATE estimate is close to the naive one. It implies the stratification may be irrelevant for the full-sample analysis. Third, we do not compute the adjusted standard error for the QTEs as it requires tuning parameters. Fourth, the QTEs provide us a new insight that the impact of supplementation on grade promotion is only significantly positive at 25% among the three quantiles. This may imply that the policy of reducing iron deficits is more effective for lower-ranked students.

Table VII. Grades Points Average

	s/adj	s/W	s/CA	ipw/W	ipw/CA
ATE	0.35	0.35	0.35	0.37	0.37
	(0.16)	(0.16)	(0.17)	(0.16)	(0.17)
QTE,25%		0.43	0.43	0.43	0.43
		(0.15)	(0.15)	(0.15)	(0.15)
QTE,50%		0.29	0.29	0.29	0.29
		(0.22)	(0.23)	(0.22)	(0.24)
QTE,75%		0.35	0.35	0.36	0.36
		(0.25)	(0.24)	(0.25)	(0.25)

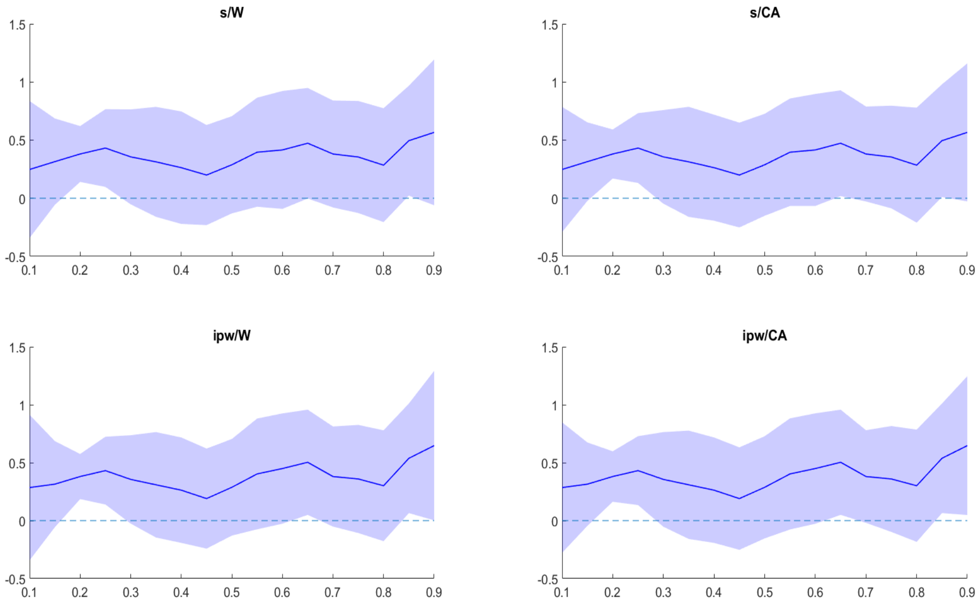


Figure 1.2. 95% Point-wise Confidence Interval for Quantile Treatment Effects

In order to provide more details on the QTE estimates, we plot the 95% point-wise confidence band in Figure 1.2 with quantile index ranging from 0.1 to 0.9. The solid line and the shadow area represent the point estimate and its 95% point-wise confidence interval, respectively. The confidence interval is constructed by

$$[\hat{\beta} - 1.96\hat{\sigma}(\hat{\beta}), \hat{\beta} + 1.96\hat{\sigma}(\hat{\beta})],$$

where $\hat{\beta}$ and $\hat{\sigma}(\hat{\beta})$ are the point estimates and the corresponding standard errors described above. As we expected, all the four findings look the same and the estimates are only significantly positive at low quantiles (15%–30%).

1.8.4 Subsample Results

Following Chong et al. (2016), we further split the sample into two based on whether the student is anemic, i.e., $Anem_i = 0$ or 1. We anticipate that there is no treatment effect for the nonanemic students and positive effects for anemic ones. In this subsample analysis, the covariate-adaptive bootstrap is infeasible, as in each sub-group, the strong-balance condition may be lost and the treatment assignment rule is not necessarily SBR and is generally unknown.[9] However, the weighted bootstrap is still feasible as it does not require the knowledge of the treatment assignment rule. According to Theorem 1.4.1, the IPW estimator paired with the weighted bootstrap standard error is valid if

$$\sup_{s \in \mathcal{S}} \left| \frac{D_n^{(1)}(s)}{n^{(1)}(s)} \right| \equiv \sup_{s \in \mathcal{S}} \left| \frac{\sum_{i=1}^n (A_i - \pi) 1\{S_i = s\} 1\{Anem_i = 1\}}{\sum_{i=1}^n 1\{S_i = s\} 1\{Anem_i = 1\}} \right| = o_p(1) \tag{1.8.1}$$

and

$$\sup_{s \in \mathcal{S}} \left| \frac{D_n^{(0)}(s)}{n^{(0)}(s)} \right| \equiv \sup_{s \in \mathcal{S}} \left| \frac{\sum_{i=1}^n (A_i - \pi) 1\{S_i = s\} 1\{Anem_i = 0\}}{\sum_{i=1}^n 1\{S_i = s\} 1\{Anem_i = 0\}} \right| = o_p(1). \tag{1.8.2}$$

We maintain this mild condition in this section. In our sample,

$$\sup_{s \in \mathcal{S}} \left| \frac{D_n^{(1)}(s)}{n^{(1)}(s)} \right| = 0 \quad \text{and} \quad \sup_{s \in \mathcal{S}} \left| \frac{D_n^{(0)}(s)}{n^{(0)}(s)} \right| = 0.071,$$

which indicate that (1.8.1) and (1.8.2) are plausible.

From Table VIII and Figure 1.3, we see that the QTE estimates are significantly positive for the anemic students when the quantile index is between around 20%–75%, but are insignificant for nonanemic students. The lack of significance at very low and high quantiles for the anemic subsample may be due to a poor asymptotic normal approximation at extreme quantiles. To extend the inference of extremal QTEs in Zhang (2018) to the context of covariate-adaptive randomization is an interesting topic for future research. We also note that for both subsamples, the weighted bootstrap standard errors for the SQR estimators are larger than those for the IPW estimators, which is consistent with Theorem 1.4.1. It implies, for both sub-groups, the stratification is relevant.

[9] As the anonymous referee pointed out, it is possible to implement the covariate-adaptive bootstrap on the full sample and pick out the observations in the subsample to construct a bootstrap subsample. The analysis can then be repeated on this covariate-adaptive bootstrap subsample. Establishing the validity of this procedure is left as a topic for future research.

Table VIII. Grades Points Average for Subsamples

	Anemic		Nonanemic	
	s/W	ipw/W	s/W	ipw/W
ATE	0.67	0.69	0.13	0.19
	(0.23)	(0.20)	(0.23)	(0.20)
QTE, 25%	0.74	0.76	0.14	0.22
	(0.24)	(0.22)	(0.28)	(0.26)
QTE, 50%	1.05	1.05	-0.14	-0.14
	(0.29)	(0.27)	(0.29)	(0.27)
QTE, 75%	0.71	0.76	0.14	0.14
	(0.36)	(0.32)	(0.39)	(0.37)

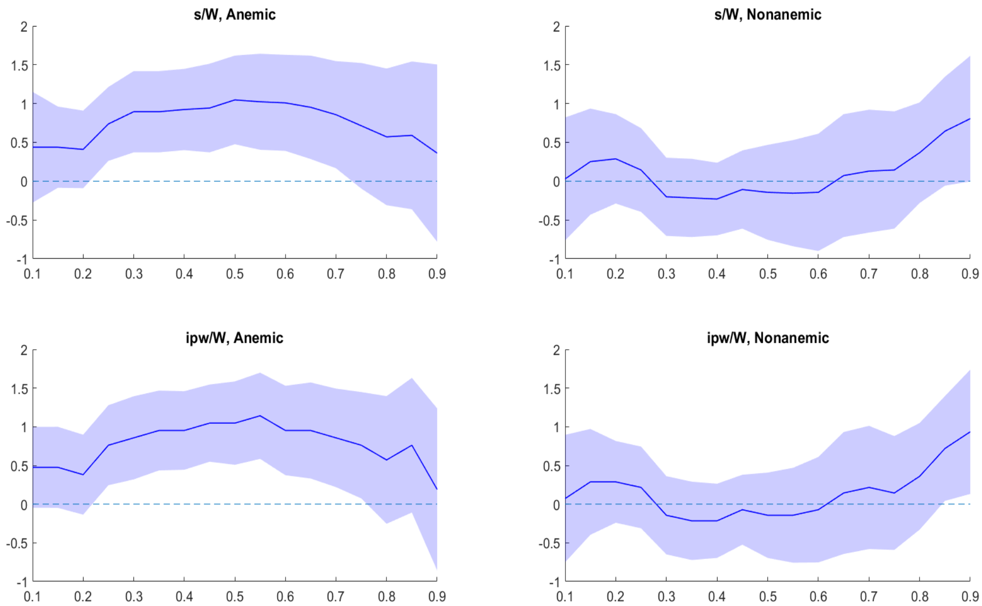


Figure 1.3. 95% Point-wise Confidence Interval for Anemic and Nonanemic Students

1.9 Conclusion

This paper studies the estimation and bootstrap inference for QTEs under covariate-adaptive randomization. We show that the weighted bootstrap standard error is only valid for the IPW estimator while the covariate-adaptive bootstrap standard error is valid for both SQR and IPW estimators. In the empirical application, we find that the QTE of iron supplementation on grade promotion is trivial for nonanemic students, while the impact is significantly positive for middle-ranked anemic students.

1.10 Appendix A

1.10.1 Proof of Theorem 1.3.1

Let $u = (u_0, u_1)' \in \Re^2$ and

$$L_n(u, \tau) = \sum_{i=1}^{n} \left[\rho_\tau(Y_i - \dot{A}_i'\beta(\tau) - \dot{A}_i'u/\sqrt{n}) - \rho_\tau(Y_i - \dot{A}_i'\beta(\tau)) \right].$$

Then, by the change of variable, we have that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = \arg\min_u L_n(u, \tau).$$

Notice that $L_n(u, \tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we aim to show that there exists

$$g_n(u, \tau) = -u'W_n(\tau) + \frac{1}{2}u'Q(\tau)u$$

such that (1) for each u,

$$\sup_{\tau \in \Upsilon} |L_n(u, \tau) - g_n(u, \tau)| \xrightarrow{p} 0;$$

(2) the maximum eigenvalue of $Q(\tau)$ is bounded from above and the minimum eigenvalue of $Q(\tau)$ is bounded away from 0, uniformly over $\tau \in \Upsilon$; (3) $W_n(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau)$ uniformly over $\tau \in \Upsilon$, in which $\tilde{\mathcal{B}}(\cdot)$ is some Gaussian process. Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) = [Q(\tau)]^{-1}W_n(\tau) + r_n(\tau),$$

where $\sup_{\tau \in \Upsilon} ||r_n(\tau)|| = o_p(1)$. In addition, by (3), we have, uniformly over $\tau \in \Upsilon$,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightsquigarrow [Q(\tau)]^{-1}\tilde{\mathcal{B}}(\tau) \equiv \mathcal{B}(\tau).$$

The second element of $\mathcal{B}(\tau)$ is $\mathcal{B}_{sqr}(\tau)$ stated in Theorem 1.3.1. In the following, we prove requirements (1)–(3) in three steps.

Step 1. By Knight's identity (Knight (1998)), we have

$$
\begin{aligned}
&L_n(u,\tau)\\
=&-u'\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\dot{A}_i\left(\tau-1\{Y_i\le \dot{A}_i'\beta(\tau)\}\right)+\sum_{i=1}^{n}\int_0^{\frac{\dot{A}_iu}{\sqrt{n}}}\left(1\{Y_i-\dot{A}_i'\beta(\tau)\le v\}-1\{Y_i-\dot{A}_i'\beta(\tau)\le 0\}\right)dv\\
\equiv&-u'W_n(\tau)+Q_n(u,\tau),
\end{aligned}
$$

where

$$
W_n(\tau)=\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\dot{A}_i\left(\tau-1\{Y_i\le \dot{A}_i'\beta(\tau)\}\right)
$$

and

$$
\begin{aligned}
Q_n(u,\tau)=&\sum_{i=1}^{n}\int_0^{\frac{\dot{A}_i'u}{\sqrt{n}}}\left(1\{Y_i-\dot{A}_i'\beta(\tau)\le v\}-1\{Y_i-\dot{A}_i'\beta(\tau)\le 0\}\right)dv\\
=&\sum_{i=1}^{n}A_i\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i(1)-q_1(\tau)\le v\}-1\{Y_i(1)-q_1(\tau)\le 0\}\right)dv\\
&+\sum_{i=1}^{n}(1-A_i)\int_0^{\frac{u_0}{\sqrt{n}}}\left(1\{Y_i(0)-q_0(\tau)\le v\}-1\{Y_i(0)-q_0(\tau)\le 0\}\right)dv\\
\equiv&Q_{n,1}(u,\tau)+Q_{n,0}(u,\tau).
\end{aligned}
$$

We first consider $Q_{n,1}(u,\tau)$. Following Bugni et al. (2018), we define $\{(Y_i^s(1),Y_i^s(0)):1\le i\le n\}$ as a sequence of i.i.d. random variables with marginal distributions equal to the distribution of $(Y_i(1),Y_i(0))|S_i=s$. The distribution of $Q_{n,1}(u,\tau)$ is the same as the counterpart with units ordered by strata and then ordered by $A_i=1$ first and $A_i=0$ second within each stratum, i.e.,

$$
\begin{aligned}
Q_{n,1}(u,\tau)\stackrel{d}{=}&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\int_0^{\frac{u_0+u_1}{\sqrt{n}}}\left(1\{Y_i^s(1)-q_1(\tau)\le v\}-1\{Y_i^s(1)-q_1(\tau)\le 0\}\right)dv\\
=&\sum_{s\in\mathcal{S}}\left[\Gamma_n^s(N(s)+n_1(s),\tau)-\Gamma_n^s(N(s),\tau)\right], \qquad (1.10.1)
\end{aligned}
$$

where $N(s) = \sum_{i=1}^n 1\{S_i < s\}$, $n_1(s) = \sum_{i=1}^n 1\{S_i = s\}A_i$, and

$$\Gamma_n^s(k,\tau) = \sum_{i=1}^k \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \Big(1\{Y_i^s(1) - q_1(\tau) \leq v\} - 1\{Y_i^s(1) - q_1(\tau) \leq 0\}\Big) dv.$$

In addition, note that

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in(0,1),\tau\in\Upsilon} |\Gamma_n^s(\lfloor nt \rfloor,\tau) - \mathbb{E}\Gamma_n^s(\lfloor nt \rfloor,\tau)| > \varepsilon) \\
=&\mathbb{P}(\max_{1\leq k\leq n}\sup_{\tau\in\Upsilon} |\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)| > \varepsilon) \\
\leq&3\max_{1\leq k\leq n}\mathbb{P}(\sup_{\tau\in\Upsilon} |\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)| > \varepsilon/3) \\
\leq&9\mathbb{P}(\sup_{\tau\in\Upsilon} |\Gamma_n^s(n,\tau) - \mathbb{E}\Gamma_n^s(n,\tau)| > \varepsilon/30) \\
\leq&\frac{270\mathbb{E}\sup_{\tau\in\Upsilon} |\Gamma_n^s(n,\tau) - \mathbb{E}\Gamma_n^s(n,\tau)|}{\varepsilon} = o(1).
\end{aligned} \tag{1.10.2}$$

The first inequality holds due to Lemma 1.10.1 with $S_k = \Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)$ and $||S_k|| = \sup_{\tau\in\Upsilon} |\Gamma_n^s(k,\tau) - \mathbb{E}\Gamma_n^s(k,\tau)|$. The second inequality holds due to Montgomery-Smith (1993, Theorem 1). To derive the last equality of (1.10.2), we consider the class of functions

$$\mathcal{F} = \left\{ \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \Big(1\{Y_i^s(1) - q_1(\tau) \leq v\} - 1\{Y_i^s(1) - q_1(\tau) \leq 0\}\Big) dv : \tau \in \Upsilon \right\}$$

with envelope $\frac{|u_0+u_1|}{\sqrt{n}}$ and

$$\sup_{f\in\mathcal{F}} \mathbb{E}f^2 \leq \sup_{\tau\in\Upsilon} \mathbb{E}\left[\frac{u_0+u_1}{\sqrt{n}} 1\left\{|Y_i^s(1) - q_1(\tau)| \leq \frac{u_0+u_1}{\sqrt{n}}\right\}\right]^2 \lesssim n^{-3/2}.$$

Note that $\mathcal{F}$ is a VC-class with a fixed VC index. Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$\mathbb{E}\sup_{\tau\in\Upsilon} |\Gamma_n^s(n,\tau) - \mathbb{E}\Gamma_n^s(n,\tau)| = n||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} \lesssim n\left[\sqrt{\frac{\log(n)}{n^{5/2}}} + \frac{\log(n)}{n^{3/2}}\right] = o(1).$$

Then, (1.10.2) implies that

$$\sup_{\tau\in\Upsilon}\left| Q_{n,1}(u,\tau) - \sum_{s\in\mathcal{S}} \mathbb{E}\Big[\Gamma_n^s(\lfloor n(N(s)/n + n_1(s)/n)\rfloor,\tau) - \Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\Big]\right| = o_p(1),$$

where following the convention in the empirical process literature,

$$\mathbb{E}\bigg[\Gamma_n^s(\lfloor n(N(s)/n+n_1(s)/n)\rfloor,\tau)-\Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\bigg]$$

is interpreted as

$$\mathbb{E}\bigg[\Gamma_n^s(\lfloor nt_2\rfloor,\tau)-\Gamma_n^s(\lfloor nt_1\rfloor,\tau)\bigg]_{t_2=\frac{N(s)}{n},t_2=\frac{N(s)+n_1(s)}{n}}.$$

In addition, $N(s)/n \xrightarrow{p} F(s)=F(S_i<s)$ and $n_1(s)/n \xrightarrow{p} \pi p(s)$. Thus, uniformly over $\tau\in\Upsilon$,

$$\begin{aligned}
&\mathbb{E}\bigg[\Gamma_n^s(\lfloor n(N(s)/n+n_1(s)/n)\rfloor,\tau)-\Gamma_n^s(\lfloor n(N(s)/n)\rfloor,\tau)\bigg]\\
=&n_1(s)\int_0^{\frac{u_0+u_1}{\sqrt{n}}}(F_1(q_1(\tau)+v|s)-F_1(q_1(\tau)|s))dv\\
\xrightarrow{p}&\frac{\pi p(s)f_1(q_1(\tau)|s)(u_0+u_1)^2}{2},
\end{aligned}$$

where $F_1(\cdot|s)$ and $f_1(\cdot|s)$ are the conditional CDF and PDF of Y_1 given $S=s$, respectively. Then, uniformly over $\tau\in\Upsilon$,

$$Q_{n,1}(u,\tau)\xrightarrow{p}\sum_{s\in\mathcal{S}}\frac{\pi p(s)f_1(q_1(\tau)|s)(u_0+u_1)^2}{2}=\frac{\pi f_1(q_1(\tau))(u_0+u_1)^2}{2}.$$

Similarly, we can show that, uniformly over $\tau\in\Upsilon$,

$$Q_{n,0}(u,\tau)\xrightarrow{p}\frac{(1-\pi)f_0(q_0(\tau))u_0^2}{2},$$

and thus

$$Q_n(u,\tau)\xrightarrow{p}\frac{1}{2}u'Q(\tau)u,$$

where

$$Q(\tau)=\begin{pmatrix}\pi f_1(q_1(\tau))+(1-\pi)f_0(q_0(\tau)) & \pi f_1(q_1(\tau))\\ \pi f_1(q_1(\tau)) & \pi f_1(q_1(\tau))\end{pmatrix}. \tag{1.10.3}$$

Then,

$$\sup_{\tau\in\Upsilon}|L_n(u,\tau)-g_n(u,\tau)| = \sup_{\tau\in\Upsilon}|Q_n(u,\tau)-\frac{1}{2}u'Q(\tau)u| = o_p(1).$$

This concludes the first step.

Step 2. Note that $\det(Q(\tau)) = \pi(1-\pi)f_1(q_1(\tau))f_0(q_0(\tau))$, which is bounded and bounded away from zero. In addition, it can be shown that the two eigenvalues of Q are nonnegative. This leads to the desired result.

Step 3. Let $e_1 = (1,1)'$ and $e_0 = (1,0)'$. Then, we have

$$\begin{aligned} W_n(\tau) =& e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}A_i 1\{S_i = s\}(\tau - 1\{Y_i(1)\le q_1(\tau)\}) \\ & + e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(1-A_i)1\{S_i = s\}(\tau - 1\{Y_i(0)\le q_0(\tau)\}). \end{aligned}$$

Let $m_j(s,\tau) = \mathbb{E}(\tau - 1\{Y_i(j)\le q_j(\tau)\}|S_i = s)$ and $\eta_{i,j}(s,\tau) = (\tau - 1\{Y_i(j)\le q_j(\tau)\}) - m_j(s,\tau)$, $j = 0,1$. Then,

$$\begin{aligned} W_n(\tau) =& \left[e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)\right] \\ & + \left[e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(A_i-\pi)1\{S_i = s\}m_1(s,\tau) - e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(A_i-\pi)1\{S_i = s\}m_0(s,\tau)\right] \\ & + \left[e_1 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\pi 1\{S_i = s\}m_1(s,\tau) + e_0 \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}(1-\pi)1\{S_i = s\}m_0(s,\tau)\right] \\ \equiv & W_{n,1}(\tau) + W_{n,2}(\tau) + W_{n,3}(\tau). \end{aligned} \tag{1.10.4}$$

By Lemma 1.10.2, uniformly over $\tau\in\Upsilon$,

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1,\tau_2)$, $\Sigma_2(\tau_1,\tau_2)$, and $\Sigma_3(\tau_1,\tau_2)$, respectively. Therefore, uniformly over $\tau\in\Upsilon$,

$$W_n(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau),$$

where $\tilde{\mathcal{B}}(\tau)$ is a two-dimensional Gaussian process with covariance kernel

$$\tilde{\Sigma}(\tau_1,\tau_2) = \sum_{j=1}^{3} \Sigma_j(\tau_1,\tau_2).$$

Consequently,

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightsquigarrow [Q(\tau)]^{-1}\tilde{\mathcal{B}}(\tau) \equiv \mathcal{B}(\tau),$$

where $\mathcal{B}(\tau)$ is a two-dimensional Gaussian process with covariance kernel

$$\begin{aligned}
\Sigma(\tau_1,\tau_2) =& [Q(\tau_1)]^{-1}\tilde{\Sigma}(\tau_1,\tau_2)[Q(\tau_2)]^{-1} \\
=& \frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)]\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&+ \frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)]\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\
&+ \sum_{s\in\mathcal{S}} p(s)\gamma(s)\Bigg[\frac{m_1(s,\tau_1)m_1(s,\tau_2)}{\pi^2 f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - \frac{m_1(s,\tau_1)m_0(s,\tau_2)}{\pi(1-\pi)f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \\
&- \frac{m_0(s,\tau_1)m_1(s,\tau_2)}{\pi(1-\pi)f_0(q_0(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} + \frac{m_0(s,\tau_1)m_0(s,\tau_2)}{(1-\pi)^2 f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\Bigg] \\
&+ \frac{\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \frac{\mathbb{E}m_1(S,\tau_1)m_0(S,\tau_2)}{f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \\
&+ \frac{\mathbb{E}m_0(S,\tau_1)m_1(S,\tau_2)}{f_0(q_0(\tau_1))f_1(q_1(\tau_2))}\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} + \frac{\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.
\end{aligned}$$

Focusing on the $(2,2)$-element of $\Sigma(\tau_1,\tau_2)$, we can conclude that

$$\sqrt{n}(\hat{\beta}_1(\tau) - q(\tau)) \rightsquigarrow \mathcal{B}_{sqr}(\tau),$$

where the Gaussian process $\mathcal{B}_{sqr}(\tau)$ has a covariance kernel

$$\begin{aligned}
&\Sigma_{sqr}(\tau_1,\tau_2) \\
=& \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))} \\
&+ \mathbb{E}\gamma(S)\Bigg[\frac{m_1(S,\tau_1)m_1(S,\tau_2)}{\pi^2 f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{m_1(S,\tau_1)m_0(S,\tau_2)}{\pi(1-\pi)f_1(q_1(\tau_1))f_0(q_0(\tau_2))}
\end{aligned}$$

$$+\frac{m_0(S,\tau_1)m_1(S,\tau_2)}{\pi(1-\pi)f_0(q_0(\tau_1))f_1(q_1(\tau_2))}+\frac{m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)^2f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\Bigg]$$
$$+\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right].$$

1.10.2 Proof of Theorem 1.3.2

By Knight's identity, we have

$$\sqrt{n}(\hat{q}_1(\tau)-q_1(\tau))=\arg\min_u L_n(u,\tau),$$

where

$$\begin{aligned}L_n(u,\tau)\equiv&\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}\left[\rho_\tau(Y_i-q_1(\tau)-\frac{u}{\sqrt{n}})-\rho_\tau(Y_i-q_1(\tau))\right]\\=&-L_{1,n}(\tau)u+L_{2,n}(u,\tau),\end{aligned}$$

$$L_{1,n}(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}(\tau-1\{Y_i\le q_1(\tau)\})$$

and

$$L_{2,n}(u,\tau)=\sum_{i=1}^n\frac{A_i}{\hat{\pi}(S_i)}\int_0^{\frac{u}{\sqrt{n}}}(1\{Y_i\le q_1(\tau)+v\}-1\{Y_i\le q_1(\tau)\})dv.$$

We aim to show that there exists

$$g_{ipw,n}(u,\tau)=-W_{ipw,n}(\tau)u+\frac{1}{2}Q_{ipw}(\tau)u^2 \tag{1.10.5}$$

such that (1) for each u,

$$\sup_{\tau\in\Upsilon}|L_n(u,\tau)-g_{ipw,n}(u,\tau)|\xrightarrow{p}0;$$

(2) $Q_{ipw}(\tau)$ is bounded and bounded away from zero uniformly over $\tau\in\Upsilon$. In addition, as a corollary of claim (3) below, $\sup_{\tau\in\Upsilon}|W_{ipw,1,n}(\tau)|=O_p(1)$. Therefore, by Kato (2009, Theorme 2), we have

$$\sqrt{n}(\hat{q}_1(\tau)-q_1(\tau))=Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau)+R_{ipw,1,n}(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_{ipw,1,n}(\tau)| = o_p(1)$. Similarly, we can show that

$$\sqrt{n}(\hat{q}_0(\tau) - q_0(\tau)) = Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) + R_{ipw,0,n}(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_{ipw,0,n}(\tau)| = o_p(1)$. Then,

$$\sqrt{n}(\hat{q}(\tau) - q(\tau)) = Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau) - Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) + R_{ipw,1,n}(\tau) - R_{ipw,0,n}(\tau).$$

Last, we aim to show that, (3) uniformly over $\tau \in \Upsilon$,

$$Q^{-1}_{ipw,1}(\tau)W_{ipw,1,n}(\tau) - Q^{-1}_{ipw,0}(\tau)W_{ipw,0,n}(\tau) \rightsquigarrow \mathcal{B}_{ipw}(\tau),$$

where $\mathcal{B}_{ipw}(\tau)$ is a scalar Gaussian process with covariance kernel $\Sigma_{ipw}(\tau_1,\tau_2)$. We prove claims (1)–(3) in three steps.

Step 1. For $L_{1,n}(\tau)$, we have

$$\begin{aligned}
L_{1,n}(\tau) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{s\in\mathcal{S}}\frac{A_i}{\pi}1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
&- \sum_{i=1}^n\sum_{s\in\mathcal{S}}\frac{A_i 1\{S_i = s\}(\hat{\pi}(s) - \pi)}{\sqrt{n}\hat{\pi}(s)\pi}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{s\in\mathcal{S}}\frac{A_i}{\pi}1\{S_i = s\}(\tau - 1\{Y_i(1) \le q_1(\tau)\}) \\
&- \sum_{i=1}^n\sum_{s\in\mathcal{S}}\frac{A_i 1\{S_i = s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}D_n(s) - \sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{\sqrt{n}\hat{\pi}(s)} \\
=& \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}\pi}m_1(s,\tau) + \sum_{i=1}^n\frac{m_1(S_i,\tau)}{\sqrt{n}} \\
&- \sum_{i=1}^n\sum_{s\in\mathcal{S}}\frac{A_i 1\{S_i = s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}D_n(s) - \sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{\sqrt{n}\hat{\pi}(s)} \\
=& W_{ipw,1,n}(\tau) + R_{ipw}(\tau),
\end{aligned}$$

where

$$W_{ipw,1,n}(\tau) = \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{i=1}^n\frac{m_1(S_i,\tau)}{\sqrt{n}} \tag{1.10.6}$$

and

$$
\begin{aligned}
&R_{ipw}(\tau)\\
=&-\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\frac{A_i1\{S_i=s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau)-\sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}D_n(s)+\sum_{s\in\mathcal{S}}\frac{D_n(s)m_1(s,\tau)}{\sqrt{n}}\left(\frac{1}{\pi}-\frac{1}{\hat{\pi}(s)}\right)\\
=&-\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\frac{A_i1\{S_i=s\}D_n(s)}{n(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau),
\end{aligned}
$$

where we use the fact that $\hat{\pi}(s)-\pi=\frac{D_n(s)}{n(s)}$. By the same argument in Claim (1) of the proof of Lemma 1.10.2, we have, for every $s\in\mathcal{S}$,

$$
\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}A_i1\{S_i=s\}\eta_{i,1}(s,\tau)\right|\stackrel{d}{=}\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=N(s)+1}^{N(s)+n(s)}\tilde{\eta}_{i,1}(s,\tau)\right|=O_p(1), \qquad (1.10.7)
$$

where $\tilde{\eta}_{i,j}(s,\tau)=\tau-1\{Y_i^s(j)\leq q_j(\tau)\}-m_j(s,\tau)$, for $j=0,1$, where $\{Y_i^s(0),Y_i^s(1)\}_{i\geq 1}$ are the same as defined in Step 1 in the proof of Theorem 1.3.1.

Because of (1.10.7) and the fact that $\frac{D_n(s)}{n(s)}=o_p(1)$, we have

$$
\sup_{\tau\in\Upsilon}|R_{ipw}(\tau)|=o_p(1).
$$

For $L_{2,n}(u,\tau)$, we have

$$
\begin{aligned}
L_{2,n}(u,\tau)=&\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\int_0^{\frac{u}{\sqrt{n}}}(1\{Y_i^s(1)\leq q_1(\tau)+v\}-1\{Y_i^s(1)\leq q_1(\tau)+v\})dv\\
=&\sum_{s\in\mathcal{S}}\frac{1}{\hat{\pi}(s)}\left[\Gamma_n^s(N(s)+n_1(s),\tau)-\Gamma_n^s(N(s),\tau)\right],
\end{aligned}
$$

where

$$
\Gamma_n^s(k,\tau)=\sum_{i=1}^{k}\int_0^{\frac{u}{\sqrt{n}}}(1\{Y_i^s(1)\leq q_1(\tau)+v\}-1\{Y_i^s(1)\leq q_1(\tau)+v\})dv.
$$

By the same argument in (1.10.2), we can show that

$$
\sup_{t\in(0,1),\tau\in\Upsilon}|\Gamma_n^s(\lfloor nt\rfloor,\tau)-\mathbb{E}\Gamma_n^s(\lfloor nt\rfloor,\tau)|=o_p(1).
$$

In addition,

$$\mathbb{E}\Gamma_n^s(N(s)+n_1(s),\tau) - \mathbb{E}\Gamma_n^s(N(s),\tau) \xrightarrow{p} \frac{\pi p(s) f_1(q_1(\tau)|s)u^2}{2}.$$

Therefore,

$$\sup_{\tau\in\Upsilon}\left|L_{2,n}(u,\tau) - \frac{f_1(q_1(\tau))u^2}{2}\right| = o_p(1),$$

where we use the fact that $\hat{\pi}(s) - \pi = \frac{D_n(s)}{n(s)} = o_p(1)$ and

$$\sum_{s\in\mathcal{S}} p(s) f_1(q_1(\tau)|s) = f_1(q_1(\tau)).$$

This establishes (1.10.5) with $Q_{ipw,1}(\tau) = f_1(q_1(\tau))$ and $W_{ipw,n}(\tau)$ defined in (1.10.6).

Step 2. Statement (2) holds by Assumption 2.

Step 3. By a similar argument in Step 1, we have

$$W_{ipw,0,n}(\tau) = \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{(1-A_i)1\{S_i = s\}}{1-\pi}\eta_{i,0}(s,\tau) + \sum_{i=1}^n \frac{m_0(S_i,\tau)}{\sqrt{n}}$$

and $Q_{ipw,0}(\tau) = f_0(q_0(\tau))$. Therefore,

$$\begin{aligned}\sqrt{n}(\hat{q}-q) =& \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] \\ &+ \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right] + R_{ipw,n}(\tau) \\ =& \mathcal{W}_{n,1}(\tau) + \mathcal{W}_{n,2}(\tau) + R_{ipw,n}(\tau)\end{aligned} \tag{1.10.8}$$

where $\sup_{\tau\in\Upsilon}|R_{ipw,n}(\tau)| = o_p(1)$. Last, Lemma 1.10.3 establishes that

$$(\mathcal{W}_{n,1}(\tau), \mathcal{W}_{n,2}(\tau)) \rightsquigarrow (\mathcal{B}_{ipw,1}(\tau), \mathcal{B}_{ipw,2}(\tau)),$$

where $(\mathcal{B}_{ipw,1}(\tau), \mathcal{B}_{ipw,2}(\tau))$ are two mutually independent scalar Gaussian processes with covariance kernels

$$\Sigma_{ipw,1}(\tau_1,\tau_2) = \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}$$

and

$$\Sigma_{ipw,2}(\tau_1,\tau_2) = \mathbb{E}\left(\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right)\left(\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right),$$

respectively. In particular, the asymptotic variance for $\hat{q}$ is

$$\zeta_Y^2(\pi,\tau) + \zeta_S^2(\tau),$$

where $\zeta_Y^2(\pi,\tau)$ and $\zeta_S^2(\tau)$ are the same as those in the proof of Theorem 1.3.1.

1.10.3 Proof of Theorem 1.4.1

First, we consider the weighted bootstrap for the SQR estimator. Note that

$$\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau)) = \arg\min_u L_n^w(u,\tau),$$

where

$$L_n^w(u,\tau) = \sum_{i=1}^n \xi_i \left[\rho_\tau(Y_i - \dot{A}_i'\beta(\tau) - \dot{A}_i'u/\sqrt{n}) - \rho_\tau(Y_i - \dot{A}_i'\beta(\tau))\right].$$

Similar to the proof of Theorem 1.3.1, we can show that

$$\sup_{\tau\in\Upsilon} |L_n^w(u,\tau) - g_n^w(u,\tau)| \to 0,$$

where

$$g_n^w(u,\tau) = -u'W_n^w(\tau) + \frac{1}{2}u'Q(\tau)u,$$

$$W_n^w(\tau) = \sum_{i=1}^n \frac{\xi_i}{\sqrt{n}}\dot{A}_i\left(\tau - 1\{Y_i \leq \dot{A}'\beta(\tau)\}\right),$$

and $Q(\tau)$ is defined in (1.10.3). Therefore, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w(\tau) - \beta(\tau)) = [Q(\tau)]^{-1}W_n^w(\tau) + r_n^w(\tau),$$

where $\sup_{\tau\in\Upsilon}||r_n^w(\tau)|| = o_p(1)$. By Theorem 1.3.1,

$$\sqrt{n}(\hat{\beta}^w(\tau) - \hat{\beta}(\tau)) = [Q(\tau)]^{-1}\sum_{i=1}^{n}\frac{\xi_i - 1}{\sqrt{n}}\dot{A}_i\left(\tau - 1\{Y_i \leq \dot{A}'\beta(\tau)\}\right) + o_p(1),$$

where the $o_p(1)$ term holds uniformly over $\tau \in \Upsilon$. In addition, Lemma 1.10.4 shows that, conditionally on data, the second element of $[Q(\tau)]^{-1}\sum_{i=1}^{n}\frac{\xi_i-1}{\sqrt{n}}\dot{A}_i\left(\tau - 1\{Y_i \leq \dot{A}'\beta(\tau)\}\right)$ converges to $\tilde{\mathcal{B}}_{sqr}(\tau)$ uniformly over $\tau \in \Upsilon$. This leads to the desired result for the weighted bootstrap simple quantile regression estimator.

Next, we turn to the IPW estimator. Denote $\hat{q}_j^w(\tau)$, $j = 0, 1$ the weighted bootstrap counterpart of $\hat{q}_j(\tau)$. We have

$$\sqrt{n}(\hat{q}_1^w(\tau) - q_1(\tau)) = \arg\min_u L_n^w(u, \tau),$$

where

$$\begin{aligned} L_n^w(u,\tau) =& \sum_{i=1}^{n}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}\left[\rho_\tau(Y_i - q_1(\tau) - \frac{u}{\sqrt{n}}) - \rho_\tau(Y_i - q_1(\tau))\right] \\ \equiv& - L_{1,n}^w(\tau)u + L_{2,n}^w(u,\tau), \end{aligned}$$

where

$$L_{1,n}^w(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}(\tau - 1\{Y_i \leq q_1(\tau)\})$$

and

$$L_{2,n}^w(\tau) = \sum_{i=1}^{n}\frac{\xi_i A_i}{\hat{\pi}^w(S_i)}\int_0^{\frac{u}{\sqrt{n}}}(1\{Y_i \leq q_1(\tau) + v\} - 1\{Y_i \leq q_1(\tau)\})dv.$$

Recall

$$D_n^w(s) = \sum_{i=1}^{n}\xi_i(A_i - \pi)1\{S_i = s\}, \quad n^w(s) = \sum_{i=1}^{n}\xi_i 1\{S_i = s\},$$

and

$$\hat{\pi}^w(s) = \frac{\sum_{i=1}^{n}\xi_i A_i 1\{S_i = s\}}{n^w(s)} = \pi + \frac{D_n^w(s)}{n^w(s)}.$$

Then, for $L^w_{1,n}(\tau)$, we have

$$
\begin{aligned}
L^w_{1,n}(\tau) =& \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\}(\hat{\pi}^w(s) - \pi)}{\sqrt{n}\hat{\pi}^w(s)\pi}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i}{\pi} 1\{S_i = s\}(\tau - 1\{Y_i(1) \leq q_1(\tau)\}) \\
& - \sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{\xi_i A_i 1\{S_i = s\} D^w_n(s)}{n^w(s)\sqrt{n}\hat{\pi}(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D^w_n(s) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{\sqrt{n}\hat{\pi}^w(s)} \\
=& \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{s \in \mathcal{S}} \frac{D^w_n(s)}{\sqrt{n}\pi} m_1(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_1(S_i,\tau)}{\sqrt{n}} \\
& - \sum_{s \in \mathcal{S}} D^w_n(s) \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D^w_n(s) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{\sqrt{n}\hat{\pi}^w(s)} \\
=& W^w_{ipw,1,n}(\tau) + R^w_{ipw}(\tau),
\end{aligned}
$$

where

$$
W^w_{ipw,1,n}(\tau) = \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{\pi}\eta_{i,1}(s,\tau) + \sum_{i=1}^n \frac{\xi_i m_1(S_i,\tau)}{\sqrt{n}} \tag{1.10.9}
$$

and

$$
\begin{aligned}
& R^w_{ipw}(\tau) \\
=& - \sum_{s \in \mathcal{S}} D^w_n(s) \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi}\eta_{i,1}(s,\tau) - \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi} D^w_n(s) + \sum_{s \in \mathcal{S}} \frac{D^w_n(s) m_1(s,\tau)}{\sqrt{n}} (\frac{1}{\pi} - \frac{1}{\hat{\pi}^w
\end{aligned}
$$

$$
= - \sum_{s \in \mathcal{S}} D^w_n(s) \sum_{i=1}^n \frac{\xi_i A_i 1\{S_i = s\}}{n^w(s)\sqrt{n}\hat{\pi}^w(s)\pi}\eta_{i,1}(s,\tau).
$$

In the following, we aim to show $D^w_n(s)/n^w(s) = o_p(1)$ and

$$
\sup_{\tau \in \Upsilon, s \in \mathcal{S}} |\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)| = O_p(\sqrt{n}).
$$

For the first claim, we note that $n^w(s)/n(s) \xrightarrow{p} 1$ and $D_n(s)/n(s) \xrightarrow{p} 0$. Therefore, we only need to show

$$\frac{D_n^w(s) - D_n(s)}{n(s)} = \sum_{i=1}^{n} \frac{(\xi_i - 1)(A_i - \pi)1\{S_i = s\}}{n(s)} \xrightarrow{p} 0.$$

As $n(s) \to \infty$ a.s., given data,

$$\begin{aligned}\frac{1}{n(s)} \sum_{i=1}^{n} (A_i - \pi)^2 1\{S_i = s\} =& \frac{1}{n} \sum_{i=1}^{n} \left(A_i - \pi - 2\pi(A_i - \pi) + \pi - \pi^2\right) 1\{S_i = s\} \\ =& \frac{D_n(s) - 2\pi D_n(s)}{n(s)} + \pi(1-\pi) \xrightarrow{p} \pi(1-\pi).\end{aligned}$$

Then, by the Lindeberg CLT, conditionally on data,

$$\frac{1}{\sqrt{n(s)}} \sum_{i=1}^{n} (\xi_i - 1)(A_i - \pi)1\{S_i = s\} \rightsquigarrow N(0, \pi(1-\pi)) = O_p(1),$$

and thus

$$\frac{D_n^w(s) - D_n(s)}{n(s)} = O_p(n^{-1/2}(s)) = o_p(1).$$

This leads to the first claim. For the second claim, we note that

$$\sum_{i=1}^{n} \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i \tilde{\eta}_{i,1}(s,\tau).$$

We can show the RHS of the above display is $O_p(\sqrt{n})$ for all $s \in \mathcal{S}$ following the same argument used in Claim (1) of the proof of Lemma 1.10.2. Given these two claims and by noticing that

$$\hat{\pi}^w(s) - \pi = \frac{D_n^w(s)}{n^w(s)} = o_p(1),$$

we have

$$\sup_{\tau \in \Upsilon} |R_{ipw}^w(\tau)| = o_p(1).$$

Similar to the argument used to derive the limit of $L_{2,n}(\tau)$ in the proof of Theorem 1.3.2, we can show that

$$\sup_{\tau\in\Upsilon}|L^w_{2,n}(u,\tau)-\frac{f_1(q_1(\tau))u^2}{2}|=o_p(1).$$

Therefore,

$$\sqrt{n}(\hat{q}^w_1(\tau)-q_1(\tau))=\frac{W^w_{ipw,1,n}(\tau)}{f_1(q_1(\tau))}+R^w_1(\tau),$$

where $\sup_{\tau\in\Upsilon}|R^w_1(\tau)|=o_p(1)$. Similarly,

$$\sqrt{n}(\hat{q}^w_0(\tau)-q_0(\tau))=\frac{W^w_{ipw,0,n}(\tau)}{f_0(q_0(\tau))}+R^w_0(\tau),$$

where

$$W^w_{ipw,0,n}(\tau)=\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\xi_i(1-A_i)1\{S_i=s\}}{1-\pi}\eta_{i,0}(s,\tau)+\sum_{i=1}^n\frac{\xi_i m_0(S_i,\tau)}{\sqrt{n}}$$

and $\sup_{\tau\in\Upsilon}|R^w_0(\tau)|=o_p(1)$. Therefore,

$$\begin{aligned}&\sqrt{n}(\hat{q}^w(\tau)-\hat{q}(\tau))\\=&\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)\Bigg\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\\&+\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\Bigg\}+o_p(1),\end{aligned}$$

where the $o_p(1)$ term holds uniformly over $\tau\in\Upsilon$. In order to show the conditional weak convergence, we only need to show the conditionally stochastic equicontinuity and finite-dimensional convergence. The former can be shown in the same manner as Lemma 1.10.4. For the latter, we note that

$$\begin{aligned}&\frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^n\Bigg\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}+\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\\=&\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\Bigg\{\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\Bigg\}^2+\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\Bigg\{\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\Bigg\}^2\\&+\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{i=1}^n\Bigg\{\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right]1\{S_i=s\}\Bigg\}^2\end{aligned}$$

$$
\begin{aligned}
&+\sum_{s\in\mathcal{S}}\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\right\}\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right] \\
&-\sum_{s\in\mathcal{S}}\frac{2}{n}\sum_{i=1}^{n}\left\{\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right\}\left[\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right] \\
&\xrightarrow{p}\zeta_Y^2(\pi,\tau)+\zeta_S^2(\tau).
\end{aligned}
$$

Note that the RHS of the above display is the same as the asymptotic variance of the original estimator $\hat{q}(\tau)$. By the CLT conditional on data, we can establish the one-dimensional weak convergence. Then, by the Cramér-Wold Theorem, we can extend such result to any finite dimension. This concludes the proof.

1.10.4 Proof of Theorem 1.5.1

It suffices to prove the theorem with $\hat{q}(\tau)$ replaced by

$$
\begin{aligned}
\tilde{q}(\tau) =& q(\tau)+\left[\sum_{s\in\mathcal{S}}\sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor}\frac{\tilde{\eta}_{i,1}(s,\tau)}{n\pi f_1(q_1(\tau))}-\sum_{s\in\mathcal{S}}\sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor}\frac{\tilde{\eta}_{i,0}(s,\tau)}{n(1-\pi)f_0(q_0(\tau))}\right] \\
&+\left[\sum_{i=1}^{n}\frac{1}{n}\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right],
\end{aligned}
$$

as we have shown in Theorem 1.3.2 that

$$
\sup_{\tau\in\Upsilon}|\tilde{q}(\tau)-\hat{q}(\tau)|=o_p(1/\sqrt{n}).
$$

We first consider the SQR estimator. Note that

$$
\sqrt{n}(\hat{\beta}^*(\tau)-\beta(\tau))=\arg\min_u L_n^*(u,\tau),
$$

where $L_n^*(u,\tau)=\sum_{i=1}^n\left[\rho_\tau(Y_i^*-\dot{A}_i^{*\prime}\beta(\tau)-\dot{A}_i^{*\prime}u/\sqrt{n})-\rho_\tau(Y_i^*-\dot{A}_i^{*\prime}\beta(\tau))\right]$. Then, $\hat{\beta}_1^*(\tau)$, the bootstrap counterpart of the SQR estimator, is just the second element of $\hat{\beta}^*(\tau)$. Similar to the proof of Theorem 1.3.1,

$$
L_n^*(u,\tau)=-u'W_n^*(\tau)+Q_n^*(u,\tau),
$$

where

$$W_n^*(\tau) = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \dot{A}_i^*(\tau - 1\{Y_i^* \leq \dot{A}_i^{*\prime} \beta(\tau)\})$$

and

$$\begin{aligned}
Q_n^*(u,\tau) =& \sum_{i=1}^{n} \int_0^{\frac{\dot{A}_i^{*\prime} u}{\sqrt{n}}} \Big(1\{Y_i^* - \dot{A}_i^{*\prime}\beta(\tau) \leq v\} - 1\{Y_i^* - \dot{A}_i^{*\prime}\beta(\tau) \leq 0\}\Big)\, dv \\
=& \sum_{i=1}^{n} A_i^* \int_0^{\frac{u_0+u_1}{\sqrt{n}}} (1\{Y_i^*(1) - q_1(\tau) \leq v\} - 1\{Y_i^*(1) - q_1(\tau) \leq 0\})\, dv \\
&+ \sum_{i=1}^{n} (1 - A_i^*) \int_0^{\frac{u_0}{\sqrt{n}}} (1\{Y_i^*(0) - q_0(\tau) \leq v\} - 1\{Y_i^*(0) - q_0(\tau) \leq 0\})\, dv \\
\equiv& Q_{n,1}^*(u,\tau) + Q_{n,0}^*(u,\tau). \qquad (1.10.10)
\end{aligned}$$

Define $\eta_{i,j}^*(s,\tau) = (\tau - 1\{Y_i^*(j) \leq q_j(\tau)\}) - m_j(s,\tau)$ and $\tilde{\eta}_{i,j}(s,\tau) = \tau - 1\{Y_i^s(j) \leq q_j(\tau)\} - m_j(s,\tau)$, $j = 0,1$, where $Y_i^s(j)$ is defined in the proof of Theorem 1.3.1. Then, we have

$$\begin{aligned}
W_n^*(\tau) =& e_1 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} A_i^* 1\{S_i^* = s\}(\tau - 1\{Y_i^*(1) \leq q_1(\tau)\}) \\
&+ e_0 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (1 - A_i^*) 1\{S_i^* = s\}(\tau - 1\{Y_i^*(0) \leq q_0(\tau)\}) \\
=& \left[e_1 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} A_i^* 1\{S_i^* = s\} \eta_{i,1}^*(s,\tau) + e_0 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (1 - A_i^*) 1\{S_i^* = s\} \eta_{i,0}^*(s,\tau) \right] \\
&+ \left[e_1 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (A_i^* - \pi) 1\{S_i^* = s\} m_1(s,\tau) - e_0 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (A_i^* - \pi) 1\{S_i^* = s\} m_0(s,\tau) \right] \\
&+ \left[e_1 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \pi 1\{S_i^* = s\} m_1(s,\tau) + e_0 \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} (1 - \pi) 1\{S_i^* = s\} m_0(s,\tau) \right] \\
\equiv& W_{n,1}^*(\tau) + W_{n,2}^*(\tau) + W_{n,3}^*(\tau).
\end{aligned}$$

By Lemma 1.10.5, there exists a sequence of independent Poisson(1) random variables

$\{\xi_i^s\}_{i\geq 1, s\in\mathcal{S}}$ such that $\{\xi_i^s\}_{i\geq 1, s\in\mathcal{S}} \perp\!\!\!\perp \{A_i^*, S_i^*, Y_i, A_i, S_i\}_{i\geq 1}$,

$$\sum_{i=1}^{n} A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s \tilde{\eta}_{i,1}(s,\tau) + R_1^*(s,\tau),$$

and

$$\sum_{i=1}^{n}(1 - A_i^*) 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau) = \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \xi_i^s \tilde{\eta}_{i,0}(s,\tau) + R_0^*(s,\tau),$$

where $\sup_{\tau\in\Upsilon}(|R_1^*(s,\tau)| + |R_0^*(s,\tau)|) = o_p(\sqrt{n(s)}) = o_p(\sqrt{n})$ for all $s \in \mathcal{S}$. Therefore,

$$(W_{n,1}^*(\tau), W_{n,2}^*(\tau), W_{n,3}^*(\tau)) \stackrel{d}{=} (\tilde{W}_{n,1}^*(\tau) + R(\tau), W_{n,2}^*(\tau), W_{n,3}^*(\tau))$$

where $\sup_{\tau\in\Upsilon} ||R(\tau)|| = o_p(1)$ and

$$\tilde{W}_{n,1}^*(\tau) = e_1 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{\xi_i^s}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{\xi_i^s}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau)$$

In addition, following the same argument in the proof of Lemma 1.10.2, we can further show that

$$\tilde{W}_{n,1}^*(\tau) = W_{n,1}^{**}(\tau) + R_n^*(\tau),$$

where $\sup_{\tau\in\Upsilon} ||R_n^*(\tau)|| = o_p(1)$ and

$$W_{n,1}^{**}(\tau) = e_1 \sum_{s\in\mathcal{S}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor} \frac{\xi_i^s}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s\in\mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \frac{\xi_i^s}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau).$$

By construction, $W_{n,1}^{**}(\tau) \perp\!\!\!\perp (W_{n,2}^*(\tau), W_{n,3}^*(\tau))$. Also note that $\{S_i^*\}_{i=1}^n$ are the nonparametric bootstrap draws based on the empirical CDF of $\{S_i\}_{i=1}^n$. Then, by van der Vaart and Wellner (1996, Section 3.6), there exists a sequence of independent Poisson(1) random variables $\{\tilde{\xi}_i\}_{i\geq 1}$ that is independent of data, $\{A_i^*\}$ and $\{\xi_i^s\}_{i\geq 1, s\in\mathcal{S}}$ such that

$$\sup_{\tau\in\Upsilon} ||W_{n,3}^*(\tau) - W_{n,3}^{**}(\tau)|| = o_p(1),$$

where

$$W_{n,3}^{**}(\tau) = e_1 \sum_{s\in\mathcal{S}} \sum_{i=1}^{n} \frac{\tilde{\xi}_i}{\sqrt{n}} \pi 1\{S_i = s\} m_1(s,\tau) + e_0 \sum_{s\in\mathcal{S}} \sum_{i=1}^{n} \frac{\tilde{\xi}_i}{\sqrt{n}} (1-\pi) 1\{S_i = s\} m_0(s,\tau)$$

By Lemma 1.10.6,

$$Q_n^*(u,\tau) \xrightarrow{p} \frac{1}{2} u' Q(\tau) u,$$

where $Q(\tau)$ is defined in (1.10.3). Then, by the same argument in the proof of Theorem 1.3.1, we have

$$\sqrt{n}(\hat{\beta}^*(\tau) - \beta(\tau)) = Q^{-1}(\tau)(W_{n,1}^{**}(\tau) + W_{n,2}^{*}(\tau) + W_{n,3}^{**}(\tau)) + R^*(\tau),$$

where $\sup_{\tau\in\Upsilon} ||R^*(\tau)|| = o_p(1)$. Focusing on the second element of $\hat{\beta}^*(\tau)$, we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}_1^*(\tau) - q(\tau)) =& \Bigg[\sum_{s\in\mathcal{S}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor} \frac{\xi_i^s \tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}\pi f_1(q_1(\tau))} - \sum_{s\in\mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \frac{\xi_i^s \tilde{\eta}_{i,0}(s,\tau)}{\sqrt{n}(1-\pi) f_0(q_0(\tau))} \Bigg] \\ &+ \left[\sum_{s\in\mathcal{S}} \frac{D_n^*(s)}{\sqrt{n}} \left(\frac{m_1(s,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(s,\tau)}{\pi f_0(q_0(\tau))} \right) \right] \\ &+ \left[\sum_{i=1}^{n} \frac{\tilde{\xi}_i}{\sqrt{n}} \left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))} \right) \right] + R_1^*(\tau), \end{aligned}$$

where $\sup_{\tau\in\Upsilon} |R_1^*(\tau)| = o_p(1)$. In addition, by definition, we have

$$\begin{aligned} \sqrt{n}(\tilde{q}(\tau) - q(\tau)) =& \Bigg[\sum_{s\in\mathcal{S}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor} \frac{\tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}\pi f_1(q_1(\tau))} - \sum_{s\in\mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \frac{\tilde{\eta}_{i,0}(s,\tau)}{\sqrt{n}(1-\pi) f_0(q_0(\tau))} \Bigg] \\ &+ \left[\sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))} \right) \right]. \end{aligned}$$

By taking difference of the two displays above, we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}_1^*(\tau) - \tilde{q}(\tau)) =& \Bigg[\sum_{s\in\mathcal{S}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor} \frac{(\xi_i^s - 1)\tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}\pi f_1(q_1(\tau))} - \sum_{s\in\mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \frac{(\xi_i^s - 1)\tilde{\eta}_{i,0}(s,\tau)}{\sqrt{n}(1-\pi) f_0(q_0(\tau))} \Bigg] \\ &+ \left[\sum_{s\in\mathcal{S}} \frac{D_n^*(s)}{\sqrt{n}} \left(\frac{m_1(s,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(s,\tau)}{\pi f_0(q_0(\tau))} \right) \right] \end{aligned}$$

$$+\left[\sum_{i=1}^{n}\frac{\tilde{\xi}_i-1}{\sqrt{n}}\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]+R_1^*(\tau). \tag{1.10.11}$$

Note that, conditionally on data, the first and third brackets on the RHS of the above display converge to Gaussian processes with covariance kernels

$$\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}$$

and

$$\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right],$$

uniformly over $\tau\in\Upsilon$, respectively. In addition, by Assumption 4(i), conditionally data (and thus $\{S_i\}_{i=1}^n$), the second bracket on the RHS of (1.10.11) converges to a Gaussian process with a covariance kernel

$$\mathbb{E}\gamma(S)\left[\frac{m_1(S,\tau_1)m_1(S,\tau_2)}{\pi^2 f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{m_1(S,\tau_1)m_0(S,\tau_2)}{\pi(1-\pi)f_1(q_1(\tau_1))f_0(q_0(\tau_2))}\right],$$

uniformly over $\tau\in\Upsilon$. Furthermore, we notice that these three Gaussian processes are independent. Therefore, we have, conditionally on data and uniformly over $\tau\in\Upsilon$,

$$\sqrt{n}(\hat{\beta}_1^*(\tau)-\tilde{q}(\tau))\rightsquigarrow\mathcal{B}_{sqr}(\tau),$$

where $\mathcal{B}_{sqr}(\tau)$ is defined in Theorem 1.3.1. This leads to the desired result for the simple quantile regression estimator.

Next, we briefly describe the derivation for the IPW estimator. Following the proof of Theorem 1.3.2, we have

$$\sqrt{n}(\hat{q}_1^*(\tau)-q_1(\tau))=\arg\min_u L_n^*(u,\tau),$$

where

$$\begin{aligned}L_n^*(u,\tau)\equiv&\sum_{i=1}^{n}\frac{A_i^*}{\hat{\pi}^*(S_i^*)}\left[\rho_\tau(Y_i^*-q_1(\tau)-\frac{u}{\sqrt{n}})-\rho_\tau(Y_i^*-q_1(\tau))\right]\\=&-L_{1,n}^*(\tau)u+L_{2,n}^*(u,\tau),\end{aligned}$$

and $\hat{\pi}^*(s) = \frac{n_1^*(s)}{n^*(s)}$. Then, we have

$$L_{1,n}^*(\tau) = W_{ipw,1,n}^*(\tau) + R_{ipw,1}^*(\tau),$$

where

$$W_{ipw,1,n}^*(\tau) = \sum_{s \in \mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau)}{\pi} + \sum_{i=1}^n \frac{m_1(S_i^*, \tau)}{\sqrt{n}},$$

and

$$R_{ipw,1}^*(\tau) = -\sum_{i=1}^n \sum_{s \in \mathcal{S}} \frac{A_i^* 1\{S_i^* = s\} D_n^*(s)}{n^*(s)\sqrt{n}\hat{\pi}^*(s)\pi} \eta_{i,1}^*(s,\tau).$$

By Lemma 1.10.5, $\sup_{\tau \in \Upsilon} |R_{ipw,1}^*(\tau)| = o_p(1)$. In addition, same as above, we can show that

$$\sup_{\tau \in \Upsilon} |W_{ipw,1,n}^*(\tau) - W_{ipw,1,n}^{**}(\tau)| = o_p(1),$$

where

$$W_{ipw,1,n}^{**}(\tau) = \sum_{s \in \mathcal{S}} \sum_{i=\lfloor nF(s) \rfloor + 1}^{\lfloor n(F(s)+\pi p(s)) \rfloor} \frac{\xi_i^s \tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}\pi} + \sum_{i=1}^n \frac{\tilde{\xi}_i m_1(S_i,\tau)}{\sqrt{n}}.$$

Similar to Lemma 1.10.6, we can show that, uniformly over $\tau \in \Upsilon$,

$$L_{2,n}^*(\tau) \xrightarrow{p} \frac{f_1(q_1(\tau))u^2}{2}.$$

Therefore,

$$\sqrt{n}(\hat{q}_1^*(\tau) - q_1(\tau)) = \frac{W_{ipw,1,n}^{**}(\tau)}{f_1(q_1(\tau))} + R_{ipw,1}^{**}(\tau),$$

where $\sup_{\tau \in \Upsilon} |R_{ipw,1}^{**}(\tau)| = o_p(1)$. Similarly, we can show

$$\sqrt{n}(\hat{q}_0^*(\tau) - q_0(\tau)) = \frac{W_{ipw,0,n}^{**}(\tau)}{f_0(q_0(\tau))} + R_{ipw,0}^{**}(\tau),$$

where $\sup_{\tau \in \Upsilon} |R^{**}_{ipw,0}(\tau)| = o_p(1)$ and

$$W^{**}_{ipw,0,n}(\tau) = \sum_{s \in \mathcal{S}} \sum_{i=\lfloor n(F(s)\pi p(s)) \rfloor+1}^{\lfloor n(F(s)+p(s)) \rfloor} \frac{\xi^s_i \tilde{\eta}_{i,0}(s,\tau)}{\sqrt{n\pi}} + \sum_{i=1}^{n} \frac{\tilde{\xi}_i m_0(S_i,\tau)}{\sqrt{n}}.$$

Therefore,

$$\sqrt{n}(\hat{q}^*(\tau) - \tilde{q}(\tau)) = \left[\sum_{s \in \mathcal{S}} \sum_{i=\lfloor nF(s) \rfloor+1}^{\lfloor n(F(s)+\pi p(s)) \rfloor} \frac{(\xi^s_i - 1)\tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n\pi} f_1(q_1(\tau))} - \sum_{s \in \mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s)) \rfloor+1}^{\lfloor n(F(s)+p(s)) \rfloor} \frac{(\xi^s_i - 1)\tilde{\eta}_{i,0}(s,\tau)}{\sqrt{n}(1-\pi) f_0(q_0(\tau))}\right]$$
$$+ \left[\sum_{i=1}^{n} \frac{\tilde{\xi}_i - 1}{\sqrt{n}} \left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right] + R^*_{ipw}(\tau),$$

where $\sup_{\tau \in \Upsilon} |R^*_{ipw}(\tau)| = o_p(1)$. Last, we can show that, conditionally on data and uniformly over $\tau \in \Upsilon$, the RHS of the above display weakly converges to the Gaussian process $\mathcal{B}_{ipw}(\tau)$, where $\mathcal{B}_{ipw}(\tau)$ is defined in Theorem 1.3.2.

1.10.5 Technical Lemmas

Lemma 1.10.1. *Let S_k be the k-th partial sum of Banach space valued independent identically distributed random variables, then*

$$\mathbb{P}(\max_{1 \le k \le n} ||S_k|| \ge \varepsilon) \le 3 \max_{1 \le k \le n} \mathbb{P}(||S_k|| \ge \varepsilon/3).$$

When S_k takes values on $\Re$, Lemma 1.10.1 is Peña et al. (2008, Exercise 2.3).

Proof. First suppose $\max_k \mathbb{P}(||S_n - S_k|| \ge 2\varepsilon/3) \le 2/3$. In addition, define

$$A_k = \{||S_k|| \ge \varepsilon, ||S_j|| < \varepsilon, 1 \le j < k\}.$$

Then,

$$\begin{aligned}
\mathbb{P}(\max_k ||S_k|| \ge \varepsilon) \le& \mathbb{P}(||S_n|| \ge \varepsilon/3) + \sum_{k=1}^{n} \mathbb{P}(||S_n|| \le \varepsilon/3, A_k) \\
\le& \mathbb{P}(||S_n|| \ge \varepsilon/3) + \sum_{k=1}^{n} \mathbb{P}(||S_n - S_k|| \ge 2\varepsilon/3)\mathbb{P}(A_k) \\
\le& \mathbb{P}(||S_n|| \ge \varepsilon/3) + \frac{2}{3}\mathbb{P}(\max_k ||S_k|| \ge \varepsilon).
\end{aligned}$$

This implies,

$$\mathbb{P}(\max_k ||S_k|| \geq \varepsilon) \leq 3\mathbb{P}(||S_n|| \geq \varepsilon/3).$$

On the other hand, if $\max_k \mathbb{P}(||S_n - S_k|| \geq 2\varepsilon/3) > 2/3$, then there exists k_0 such that $\mathbb{P}(||S_n - S_{k_0}|| \geq 2\varepsilon/3) > 2/3$. Thus,

$$\mathbb{P}(||S_n|| \geq \varepsilon/3) + \mathbb{P}(||S_{k_0}|| \geq \varepsilon/3) \geq 2/3.$$

This implies,

$$3 \max_{1\leq k\leq n} \mathbb{P}(||S_k|| \geq \varepsilon/3) \geq 3\max(\mathbb{P}(||S_n|| \geq \varepsilon/3), \mathbb{P}(||S_{k_0}|| \geq \varepsilon/3)) \geq 1 \geq \mathbb{P}(\max_{1\leq k\leq n} ||S_k|| \geq \varepsilon).$$

This concludes the proof. □

Lemma 1.10.2. *Let $W_{n,j}(\tau)$, $j = 1,2,3$ be defined as in (1.10.4). If Assumptions in Theorem 1.3.1 hold, then uniformly over $\tau \in \Upsilon$,*

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1,\tau_2)$, $\Sigma_2(\tau_1,\tau_2)$, and $\Sigma_3(\tau_1,\tau_2)$, respectively. The expressions for the three kernels are derived in the proof below.

Proof. We follow the general argument in the proof of Bugni et al. (2018, Lemma B.2). We divide the proof into two steps. In the first step, we show that

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \stackrel{d}{=} (W^\star_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) + o_p(1),$$

where the $o_p(1)$ term holds uniformly over $\tau \in \Upsilon$, $W^\star_{n,1}(\tau) \perp\!\!\!\perp (W_{n,2}(\tau), W_{n,3}(\tau))$, and, uniformly over $\tau \in \Upsilon$,

$$W^\star_{n,1}(\tau) \rightsquigarrow \mathcal{B}_1(\tau).$$

In the second step, we show that

$$(W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$$

uniformly over $\tau \in \Upsilon$ and $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$.

Step 1. Let $\tilde{\eta}_{i,j}(s,\tau) = \tau - 1\{Y_i^s(j) \leq q_j(\tau)\} - m_j(s,\tau)$, for $j = 0, 1$, where $\{Y_i^s(0), Y_i^s(1)\}_{i \geq 1}$ are the same as defined in Step 1 in the proof of Theorem 1.3.1. In addition, denote

$$\tilde{W}_{n,1}(\tau) = e_1 \sum_{s \in \mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s \in \mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau).$$

Then, we have

$$\{W_{n,1}(\tau) | \{A_i, S_i\}_{i=1}^n\} \stackrel{d}{=} \{\tilde{W}_{n,1}(\tau) | \{A_i, S_i\}_{i=1}^n\}.$$

Because both $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$ are only functions of $\{A_i, S_i\}_{i=1}^n$, we have

$$(W_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)) \stackrel{d}{=} (\tilde{W}_{n,1}(\tau), W_{n,2}(\tau), W_{n,3}(\tau)).$$

Let

$$W_{n,1}^\star(\tau) = e_1 \sum_{s \in \mathcal{S}} \sum_{i=\lfloor nF(s) \rfloor + 1}^{\lfloor n(F(s)+\pi p(s)) \rfloor} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + e_0 \sum_{s \in \mathcal{S}} \sum_{i=\lfloor n(F(s)+\pi p(s)) \rfloor + 1}^{\lfloor n(F(s)+p(s)) \rfloor} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau).$$

Note that $W_{n,1}^\star(\tau)$ is a function of $(Y_i^s(1), Y_i^s(0))_{i \geq 1}$ only, which is independent of $\{A_i, S_i\}_{i=1}^n$ by construction. Therefore, $W_{n,1}^\star(\tau) \perp\!\!\!\perp (W_{n,2}(\tau), W_{n,3}(\tau))$.

Furthermore, note that

$$\frac{N(s)}{n} \xrightarrow{p} F(s), \quad \frac{n_1(s)}{n} \xrightarrow{p} \pi p(s), \quad \text{and} \quad \frac{n(s)}{n} \xrightarrow{p} p(s).$$

Denote $\Gamma_{n,j}(s,t,\tau) = \sum_{i=1}^{\lfloor nt \rfloor} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,j}(s,\tau)$. In order to show $\sup_{\tau \in \Upsilon} |\tilde{W}_{n,1}(\tau) - W_{n,1}^\star(\tau)| = o_p(1)$ and $W_{n,1}^\star(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$, it suffices to show that, (1) for $j = 0, 1$ and $s \in \mathcal{S}$, the stochastic processes

$$\{\Gamma_{n,j}(s,t,\tau) : t \in (0,1), \tau \in \Upsilon\}$$

in stochastically equicontinuous; and (2) $W_{n,1}^\star(\tau)$ converges to $\mathcal{B}_1(\tau)$ in finite dimension.

Claim (1). We want to bound

$$\sup |\Gamma_{n,j}(s,t_2,\tau_2) - \Gamma_{n,j}(s,t_1,\tau_1)|,$$

where supremum is taken over $0 < t_1 < t_2 < t_1 + \varepsilon < 1$ and $\tau_1 < \tau_2 < \tau_1 + \varepsilon$ such that

$\tau_1, \tau_1 + \varepsilon \in \Upsilon$. Note that,

$$
\begin{aligned}
&\sup |\Gamma_{n,j}(s,t_2,\tau_2) - \Gamma_{n,j}(s,t_1,\tau_1)| \\
\le& \sup_{0<t_1<t_2<t_1+\varepsilon<1, \tau\in\Upsilon} |\Gamma_{n,j}(s,t_2,\tau) - \Gamma_{n,j}(s,t_1,\tau)| + \sup_{t\in(0,1),\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon} |\Gamma_{n,j}(s,t,\tau_2) - \Gamma_{n,j}(s,t,\tau_1)|.
\end{aligned}
\tag{1.10.12}
$$

Let $m = \lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor \le \lfloor n\varepsilon \rfloor + 1$. Then, for an arbitrary $\delta > 0$, by taking $\varepsilon = \delta^4$, we have

$$
\begin{aligned}
&\mathbb{P}(\sup_{0<t_1<t_2<t_1+\varepsilon<1,\tau\in\Upsilon} |\Gamma_{n,j}(s,t_2,\tau) - \Gamma_{n,j}(s,t_1,\tau)| \ge \delta) \\
=&\mathbb{P}(\sup_{0<t_1<t_2<t_1+\varepsilon<1,\tau\in\Upsilon} | \sum_{i=\lfloor nt_1\rfloor+1}^{i=\lfloor nt_2\rfloor} \tilde{\eta}_{i,j}(s,\tau)| \ge \sqrt{n}\delta) \\
=&\mathbb{P}(\sup_{0<t\le\varepsilon,\tau\in\Upsilon} | \sum_{i=1}^{\lfloor nt\rfloor} \tilde{\eta}_{i,j}(s,\tau)| \ge \sqrt{n}\delta) \\
\le&\mathbb{P}(\max_{1\le k\le\lfloor n\varepsilon\rfloor} \sup_{\tau\in\Upsilon} |S_k(\tau)| \ge \sqrt{n}\delta) \\
\le&\frac{270\mathbb{E}\sup_{\tau\in\Upsilon} |\sum_{i=1}^{\lfloor n\varepsilon\rfloor} \tilde{\eta}_{i,j}(s,\tau)|}{\sqrt{n}\delta} \\
\lesssim&\frac{\sqrt{n\varepsilon}}{\sqrt{n}\delta} \lesssim \delta,
\end{aligned}
$$

where in the first inequality, $S_k(\tau) = \sum_{i=1}^{k} \tilde{\eta}_{i,j}(s,\tau)$ and the second inequality holds due to the same argument in (1.10.2). For the third inequality, denote

$$
\mathcal{F} = \{\tilde{\eta}_{i,j}(s,\tau) : \tau \in \Upsilon\}
$$

with an envelope function $F = 2$. In addition, because $\mathcal{F}$ is a VC-class with a fixed VC-index, we have

$$
J(1,\mathcal{F}) < \infty,
$$

where

$$
J(\delta,\mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\varepsilon||F||_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon,
$$

$N(\varepsilon||F||_{Q,2}, \mathcal{F}, L_2(Q))$ is the covering number, and the supremum is taken over all discrete probability measures Q. Therefore, by van der Vaart and Wellner (1996, Theorem 2.14.1)

$$\frac{270\mathbb{E}\sup_{\tau\in\Upsilon}|\sum_{i=1}^{\lfloor n\varepsilon\rfloor}\tilde{\eta}_{i,j}(s,\tau)|}{\sqrt{n}\delta} \lesssim \frac{\sqrt{\lfloor n\varepsilon\rfloor}\left[\mathbb{E}\sqrt{\lfloor n\varepsilon\rfloor}||\mathbb{P}_{\lfloor n\varepsilon\rfloor}-\mathbb{P}||_{\mathcal{F}}\right]}{\sqrt{n}\delta} \lesssim \frac{\sqrt{\lfloor n\varepsilon\rfloor}J(1,\mathcal{F})}{\sqrt{n}\delta}.$$

For the second term on the RHS of (1.10.12), by taking $\varepsilon = \delta^4$, we have

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in(0,1),\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon}|\Gamma_{n,j}(s,t,\tau_2)-\Gamma_{n,j}(s,t,\tau_1)|\geq\delta)\\
=&\mathbb{P}(\max_{1\leq k\leq n}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon}|S_k(\tau_1,\tau_2)|\geq\sqrt{n}\delta)\\
\leq&\frac{270\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon}|\sum_{i=1}^{n}(\tilde{\eta}_{i,j}(s,\tau_2)-\tilde{\eta}_{i,j}(s,\tau_1))|}{\sqrt{n}\delta} \lesssim \delta\sqrt{\log(\frac{C}{\delta^2})},
\end{aligned}$$

where in the first equality, $S_k(\tau_1,\tau_2) = \sum_{i=1}^{k}(\tilde{\eta}_{i,j}(s,\tau_2)-\tilde{\eta}_{i,j}(s,\tau_1))$ and the first inequality follows the same argument as in (1.10.2). For the last inequality, denote

$$\mathcal{F} = \{\tilde{\eta}_{i,j}(s,\tau_2)-\tilde{\eta}_{i,j}(s,\tau_1) : \tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon\}$$

with a constant envelope function $F = C$ and

$$\sigma^2 = \sup_{f\in\mathcal{F}}\mathbb{E}f^2 \in [c_1\varepsilon, c_2\varepsilon],$$

for some constant $0 < c_1 < c_2 < \infty$. Last, $\mathcal{F}$ is nested by some VC class with a fixed VC index. Therefore, by Chernozhukov et al. (2014, Corollary 5.1),

$$\begin{aligned}
&\frac{270\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\varepsilon}|\sum_{i=1}^{n}(\tilde{\eta}_{i,j}(s,\tau_2)-\tilde{\eta}_{i,j}(s,\tau_1))|}{\sqrt{n}\delta}\\
\lesssim&\frac{\sqrt{n}\mathbb{E}||\mathbb{P}_n-\mathbb{P}||_{\mathcal{F}}}{\delta} \lesssim \sqrt{\frac{\sigma^2\log(\frac{C}{\sigma})}{\delta^2}}+\frac{C\log(\frac{C}{\sigma})}{\sqrt{n}\delta} \lesssim \delta\sqrt{\log(\frac{C}{\delta^2})},
\end{aligned}$$

where the last inequality holds by letting n be sufficiently large. Note that $\delta\sqrt{\log(\frac{C}{\delta^2})} \to 0$ as $\delta \to 0$. This concludes the proof of Claim (1).

Claim (2). For a single τ, by the triangular CLT,

$$W^{\star}_{n,1}(\tau) \rightsquigarrow N(0,\Sigma_1(\tau)),$$

where $\Sigma_1(\tau) = \pi[\tau(1-\tau)-\mathbb{E}m_1^2(S,\tau)]e_1e_1'+(1-\pi)[\tau(1-\tau)-\mathbb{E}m_0^2(S,\tau)]e_0e_0'$. The convergence in finite dimension can be proved by using the Cramér-Wold device. In particular, we can show that the covariance kernel is

$$\Sigma_1(\tau_1,\tau_2) =\pi[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)]e_1e_1' \\ +(1-\pi)[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)]e_0e_0'.$$

This concludes the proof of Claim (2), and thus leads to the desired results in Step 1.

Step 2. We first consider the marginal distributions for $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$. For $W_{n,2}(\tau)$, by Assumption 1 and the fact that $m_j(s,\tau)$ is continuous in $\tau \in \Upsilon$ $j=0,1$, we have, conditionally on $\{S_i\}_{i=1}^n$,

$$W_{n,2}(\tau)=\sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}}[e_1m_1(s,\tau)-e_0m_0(s,\tau)] \rightsquigarrow \mathcal{B}_2(\tau), \tag{1.10.13}$$

where $\mathcal{B}_2(\tau)$ is a two-dimensional Gaussian process with covariance kernel

$$\Sigma_2(\tau_1,\tau_2) \\ =\sum_{s\in\mathcal{S}}p(s)\gamma(s)\Big[e_1e_1'm_1(s,\tau_1)m_1(s,\tau_2)-e_1e_0'm_1(s,\tau_1)m_0(s,\tau_2) \\ -e_0e_1'm_0(s,\tau_1)m_1(s,\tau_2)+e_0e_0'm_0(s,\tau_1)m_0(s,\tau_2)\Big].$$

For $W_{n,3}(\tau)$, by the fact that $m_j(s,\tau)$ is continuous in $\tau\in\Upsilon$ $j=0,1$, we have that, uniformly over $\tau\in\Upsilon$,

$$W_{n,3}(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^n[e_1\pi m_1(S_i,\tau)+e_0(1-\pi)m_0(S_i,\tau)] \rightsquigarrow \mathcal{B}_3(\tau), \tag{1.10.14}$$

where $\mathcal{B}_3(\tau)$ a two-dimensional Gaussian process with covariance kernel

$$\Sigma_3(\tau_1,\tau_2) =e_1e_1'\pi^2\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)+e_1e_0'\pi(1-\pi)\mathbb{E}m_1(S,\tau_1)m_0(S,\tau_2) \\ +e_0e_1'\pi(1-\pi)\mathbb{E}m_0(S,\tau_1)m_1(S,\tau_2)+e_0e_0'(1-\pi)^2\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2).$$

In addition, we note that, for any fixed τ,

$$\mathbb{P}(W_{n,2}(\tau)\le w_1, W_{n,3}(\tau)\le w_2) =\mathbb{E}\mathbb{P}(W_{n,2}(\tau)\le w_1|\{S_i\}_{i=1}^n)1\{W_{n,3}(\tau)\le w_2\} \\ =\mathbb{E}\mathbb{P}(N(0,\Sigma_2(\tau,\tau))\le w_1)1\{W_{n,3}(\tau)\le w_2\}+o(1)$$

$$=\mathbb{P}(N(0,\Sigma_3(\tau,\tau))\le w_2)\mathbb{P}(N(0,\Sigma_2(\tau,\tau))\le w_1)+o(1).$$

This implies $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$. By the Cramér-Wold device, we can show that

$$(W_{n,2}(\tau), W_{n,3}(\tau)) \rightsquigarrow (\mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$$

jointly in finite dimension, where by an abuse of notation, $\mathcal{B}_2(\tau)$ and $\mathcal{B}_3(\tau)$ have the same marginal distributions of those in (1.10.13) and (1.10.14), respectively, and $\mathcal{B}_2(\tau) \perp\!\!\!\perp \mathcal{B}_3(\tau)$. Last, because both $W_{n,2}(\tau)$ and $W_{n,3}(\tau)$ are tight marginally, so be the joint process $(W_{n,2}(\tau), W_{n,3}(\tau))$. This concludes the proof of Step 2, and thus the whole lemma. □

Lemma 1.10.3. *Let $\mathcal{W}_{n,j}(\tau)$, $j = 1, 2$ be defined as in (1.10.8). If Assumptions in Theorem 1.3.2 hold, then uniformly over $\tau \in \Upsilon$,*

$$(\mathcal{W}_{n,1}(\tau), \mathcal{W}_{n,2}(\tau)) \rightsquigarrow (\mathcal{B}_{ipw,1}(\tau), \mathcal{B}_{ipw,2}(\tau)),$$

where $(\mathcal{B}_{ipw,1}(\tau), \mathcal{B}_{ipw,2}(\tau))$ are two independent two-dimensional Gaussian processes with covariance kernels $\Sigma_{ipw,1}(\tau_1,\tau_2)$ and $\Sigma_{ipw,2}(\tau_1,\tau_2)$, respectively. The expressions for $\Sigma_{ipw,1}(\tau_1,\tau_2)$ and $\Sigma_{ipw,2}(\tau_1,\tau_2)$ are derived in the proof below.

Proof. The proofs of weak convergence and the independence between $(\mathcal{B}_{ipw,1}(\tau), \mathcal{B}_{ipw,2}(\tau))$ are similar to that in Lemma 1.10.2, and thus, are omitted. Next, we focus on deriving the covariance kernels.

First, similar to the argument in the proof of Lemma 1.10.2,

$$\mathcal{W}_{n,1}(\tau) \overset{d}{=} \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n} f_1(q_1(\tau))} \tilde{\eta}_{i,1}(s,\tau) - \sum_{s\in\mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n} f_0(q_0(\tau))} \tilde{\eta}_{i,0}(s,\tau).$$

Because $(\tilde{\eta}_{i,1}(s,\tau), \tilde{\eta}_{i,0}(s,\tau))$ are independent across i, $n_1(s)/n \xrightarrow{p} \pi p(s)$, and $(n(s)-n_1(s))/n \xrightarrow{p} (1-\pi)p(s)$, we have

$$\Sigma_{ipw,1}(\tau_1,\tau_2) = \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1)) f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi) f_0(q_0(\tau_1)) f_0(q_0(\tau_2))}.$$

Obviously,

$$\Sigma_{ipw,2}(\tau_1,\tau_2) = \mathbb{E}\left(\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right)\left(\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right),$$

□

Lemma 1.10.4. *If Assumptions 1 and 2 hold, then conditionally on data, the second element of* $[Q(\tau)]^{-1}\sum_{i=1}^{n}\frac{\xi_i-1}{\sqrt{n}}\dot{A}_i\left(\tau - 1\{Y_i \le \dot{A}'\beta(\tau)\}\right)$ *weakly converges to* $\tilde{\mathcal{B}}_{sqr}(\tau)$, *where* $\tilde{\mathcal{B}}_{sqr}(\tau)$ *is a Gaussian process with covariance kernel* $\tilde{\Sigma}_{sqr}(\cdot,\cdot)$ *defined in Theorem 1.4.1.*

Proof. We denote the second element of $[Q(\tau)]^{-1}\sum_{i=1}^{n}\frac{\xi_i-1}{\sqrt{n}}\dot{A}_i\left(\tau - 1\{Y_i \le \dot{A}'\beta(\tau)\}\right)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i - 1)\mathcal{J}_i(s,\tau),$$

where

$$\mathcal{J}_i(s,\tau) = \mathcal{J}_{i,1}(s,\tau) + \mathcal{J}_{i,2}(s,\tau) + \mathcal{J}_{i,3}(s,\tau),$$

$$\mathcal{J}_{i,1}(s,\tau) = \frac{A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))},$$

$$\mathcal{J}_{i,2}(s,\tau) = F_1(s,\tau)(A_i - \pi)1\{S_i = s\},$$

$$F_1(s,\tau) = \frac{m_1(s,\tau)}{\pi f_1(q_1(\tau))} + \frac{m_0(s,\tau)}{(1-\pi)f_0(q_0(\tau))},$$

and

$$\mathcal{J}_{i,3}(s,\tau) = \left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i = s\}.$$

In order to show the weak convergence, we only need to show (1) conditionally stochastic equicontinuity and (2) conditional convergence in finite dimension. We divide the proof into two steps accordingly.

Step 1. In order to show the conditionally stochastic equicontinuity, it suffices to show that, for any $\varepsilon > 0$, as $n \to \infty$ followed by $\delta \to 0$,

$$\mathbb{P}_\xi\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right| \ge \varepsilon\right) \xrightarrow{p} 0,$$

where $\mathbb{P}_\xi(\cdot)$ means that the probability operator is with respect to $\xi_1,\cdots,\xi_n$ and conditional on data. Note

$$
\begin{aligned}
&\mathbb{EP}_\xi\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_i(s,\tau_1)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
=&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
\leq&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_{i,1}(s,\tau_1))\right|\geq\varepsilon/3\right)\\
&+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2)-\mathcal{J}_{i,2}(s,\tau_1))\right|\geq\varepsilon/3\right)\\
&+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,3}(s,\tau_2)-\mathcal{J}_{i,3}(s,\tau_1))\right|\geq\varepsilon/3\right).
\end{aligned}
$$

Further note that

$$
\sum_{i=1}^n(\xi_i-1)\mathcal{J}_{i,1}(s,\tau)\stackrel{d}{=}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{(\xi_i-1)\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\sum_{i=n(s)+n_1(s)+1}^{N(s)+n(s)}\frac{(\xi_i-1)\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}
$$

By the same argument in Claim (1) in the proof of Lemma 1.10.2, we have

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon/3\right)\\
\leq&\frac{3\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_{i,1}(s,\tau_1))\right|}{\varepsilon}\\
\leq&\frac{3\sqrt{c_2\delta\log(\frac{C}{c_1\delta})}+\frac{3C\log(\frac{C}{c_1\delta})}{\sqrt{n}}}{\varepsilon},
\end{aligned}
$$

where C, $c_1<c_2$ are some positive constants that are independent of (n,ε,δ). By letting $n\to\infty$ followed by $\delta\to 0$, the RHS vanishes.

For $\mathcal{J}_{i,2}$, we note that $F_1(s,\tau)$ is Lipschitz in τ. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2)-\mathcal{J}_{i,2}(s,\tau_1))\right|\geq\varepsilon/3\right)\\
\leq&\sum_{s\in\mathcal{S}}\mathbb{P}\left(C\delta\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(A_i-\pi)1\{S_i=s\}\right|\geq\varepsilon/3\right)\to 0
\end{aligned}
$$

as $n \to \infty$ followed by $\delta \to 0$, where we use the fact that

$$\sup_{s \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - 1)(A_i - \pi) 1\{S_i = s\} \right| = O_p(1).$$

To see this claim, we note that, conditionally on data,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} (A_i - \pi)^2 1\{S_i = s\} =& \frac{1}{n} \sum_{i=1}^{n} \left(A_i - \pi - 2\pi(A_i - \pi) + \pi - \pi^2 \right) 1\{S_i = s\} \\
=& \frac{D_n(s) - 2\pi D_n(s)}{n} + \pi(1-\pi) \frac{n(s)}{n} \xrightarrow{p} \pi(1-\pi)p(s).
\end{aligned}$$

Then, by the Lindeberg CLT, conditionally on data,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - 1)(A_i - \pi) 1\{S_i = s\} \rightsquigarrow N(0, \pi(1-\pi)p(s)) = O_p(1).$$

Last, by the standard maximal inequality (e.g., van der Vaart and Wellner (1996, Theorem 2.14.1)) and the fact that

$$\left(\frac{m_1(s, \tau)}{f_1(q_1(\tau))} - \frac{m_0(s, \tau)}{f_0(q_0(\tau))} \right)$$

is Lipschitz in τ, we have, as $n \to \infty$ followed by $\delta \to 0$,

$$\mathbb{P} \left(\sup_{\tau_1, \tau_2 \in \Upsilon, \tau_1 < \tau_2 < \tau_1 + \delta, s \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - 1)(\mathcal{J}_{i,3}(s, \tau_2) - \mathcal{J}_{i,3}(s, \tau_1)) \right| \geq \varepsilon/3 \right) \to 0.$$

This concludes the proof of the conditionally stochastic equicontinuity.

Step 2. We focus on the one-dimension case and aim to show that, conditionally on data, for fixed $\tau \in \Upsilon$,

$$\frac{1}{\sqrt{n}} \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} (\xi_i - 1) \mathcal{J}_i(s, \tau) \rightsquigarrow \mathcal{N}(0, \tilde{\Sigma}_{sqr}(\tau, \tau)).$$

The finite-dimensional convergence can be established similarly by the Cramér-Wold device. In view of Lindeberg-Feller central limit theorem, we only need to show that (1)

$$\frac{1}{n} \sum_{i=1}^{n} [\sum_{s \in \mathcal{S}} \mathcal{J}_i(s, \tau)]^2 \xrightarrow{p} \zeta_Y^2(\pi, \tau) + \tilde{\xi}_A^2(\pi, \tau) + \xi_S^2(\pi, \tau)$$

and (2)

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2\mathbb{E}_\xi(\xi-1)^2 1\{|\sum_{s\in\mathcal{S}}(\xi_i-1)\mathcal{J}_i(s,\tau)|\geq\varepsilon\sqrt{n}\}\to 0.$$

(2) is obvious as $|\mathcal{J}_i(s,\tau)|$ is bounded and $\max_i|\xi_i-1|\lesssim\log(n)$ as ξ_i is sub-exponential. Next, we focus on (1). We have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\Bigg\{\Bigg[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\Bigg]\\
&+F_1(s,\tau)(A_i-\pi)1\{S_i=s\}+\Bigg[\Bigg(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\Bigg)1\{S_i=s\}\Bigg]\Bigg\}^2\\
\equiv&\sigma_1^2+\sigma_2^2+\sigma_3^2+2\sigma_{12}+2\sigma_{13}+2\sigma_{23},
\end{aligned}$$

where

$$\sigma_1^2=\frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\Bigg[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\Bigg]^2,$$

$$\sigma_2^2=\frac{1}{n}\sum_{s\in\mathcal{S}}F_1^2(s,\tau)\sum_{i=1}^{n}(A_i-\pi)^2 1\{S_i=s\},$$

$$\sigma_3^2=\frac{1}{n}\sum_{i=1}^{n}\Bigg[\Bigg(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\Bigg)\Bigg]^2,$$

$$\sigma_{12}=\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\Bigg[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\Bigg]F_1(s,\tau)(A_i-\pi)1\{S_i=s\},$$

$$\sigma_{13}=\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\Bigg[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\Bigg]\Bigg[\Bigg(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\Bigg)\Bigg],$$

and

$$\sigma_{23} = \sigma_{12} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}} F_1(s,\tau)(A_i-\pi)1\{S_i=s\}\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right].$$

For σ_1^2, we have

$$\begin{aligned}
\sigma_1^2 =& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))}\right] \\
\stackrel{d}{=}& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} + \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))} \\
\xrightarrow{p}& \frac{\tau(1-\tau)-\mathbb{E}m_1^s(S,\tau)}{\pi f_1^2(q_1(\tau))} + \frac{\tau(1-\tau)-\mathbb{E}m_0^s(S,\tau)}{(1-\pi)f_0^2(q_0(\tau))} = \zeta_Y^2(\pi,\tau),
\end{aligned}$$

where the second equality holds due to the rearrangement argument in Lemma 1.10.2 and the convergence in probability holds due to uniform convergence of the partial sum process.

For σ_2^2, by Assumption 1,

$$\sigma_2^2 = \frac{1}{n}\sum_{s\in\mathcal{S}} F_1^2(s,\tau)(D_n(s) - 2\pi D_n(s) + \pi(1-\pi)1\{S_i=s\}) \xrightarrow{p} \pi(1-\pi)\mathbb{E}F_1^2(S_i,\tau) = \tilde{\xi}_A^2(\pi,\tau).$$

For σ_3^2, by the law of large number,

$$\sigma_3^2 \xrightarrow{p} \mathbb{E}\left[\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]^2 = \xi_S^2(\pi,\tau).$$

For σ_{12}, we have

$$\begin{aligned}
\sigma_{12} =& \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=1}^{n}\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=1}^{n}\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \\
\stackrel{d}{=}& \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \xrightarrow{p} 0,
\end{aligned}$$

where the last convergence holds because by Lemma 1.10.2,

$$\frac{1}{n}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\eta}_{i,1}(s,\tau) \xrightarrow{p} 0, \quad \text{and} \quad \frac{1}{n}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\tilde{\eta}_{i,0}(s,\tau) \xrightarrow{p} 0.$$

By the same argument, we can show that

$$\sigma_{13} \xrightarrow{p} 0.$$

Last, for σ_{23}, by Assumption 1,

$$\sigma_{23} = \sum_{s \in \mathcal{S}} F_1(s,\tau)\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right]\frac{D_n(s)}{n} \xrightarrow{p} 0.$$

Therefore, conditionally on data,

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s \in \mathcal{S}} \mathcal{J}_i(s,\tau)]^2 \xrightarrow{p} \zeta_Y^2(\pi,\tau) + \tilde{\xi}_A^2(\pi,\tau) + \xi_S^2(\pi,\tau).$$

□

Lemma 1.10.5. *If Assumptions 1(i) and 1(ii) hold,* $\sup_{s \in \mathcal{S}} \frac{|D_n^*(s)|}{\sqrt{n^*(s)}} = O_p(1)$*,* $\sup_{s \in \mathcal{S}} \frac{|D_n(s)|}{\sqrt{n(s)}} = O_p(1)$*, and* $n(s) \to \infty$ *for all* $s \in \mathcal{S}$*, a.s., then there exists a sequence of Poisson(1) random variables* $\{\xi_i^s\}_{i \geq 1, s \in \mathcal{S}}$ *independent of* $\{A_i^*, S_i^*, Y_i, A_i, S_i\}_{i \geq 1}$ *such that*

$$\sum_{i=1}^{n} A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s \tilde{\eta}_{i,1}(s,\tau) + R_1^*(s,\tau),$$

where $\sup_{\tau \in \Upsilon, s \in \mathcal{S}} |R_1^*(s,\tau)/\sqrt{n(s)}| = o_p(1)$*. In addition,*

$$\sup_{s \in \mathcal{S}, \tau \in \Upsilon} |\sum_{i=1}^{n} A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s,\tau)|/\sqrt{n(s)} = O_p(1). \quad (1.10.15)$$

Proof. Recall $\{Y_i^s(0), Y_i^s(1)\}_{i=1}^n$ as defined in the proof of Theorem 1.3.1 and

$$\tilde{\eta}_{i,j}(s,\tau) = \tau - 1\{Y_i^s(j) \leq q_j(\tau)\} - m_j(s,\tau),$$

$j = 0, 1$. In addition, let $\Psi_n = \{\eta_{i,1}(s,\tau)\}_{i=1}^n$,

$$\mathbb{N}_n = \{n(s)/n, n_1(s)/n, n^*(s)/n, n_1^*(s)/n\}_{s \in \mathcal{S}}$$

and given $\mathbb{N}_n$, $\{M_{ni}\}_{i=1}^n$ be a sequence of random variables such that the $n_1(s) \times 1$ vector

$$M_n^1(s) = (M_{n,N(s)+1}, \cdots, M_{n,N(s)+n_1(s)})$$

and the $(n(s) - n_1(s)) \times 1$ vector

$$M_n^0(s) = (M_{n,N(s)+n_1(s)+1}, \cdots, M_{n,N(s)+n(s)})$$

satisfy:

1. $M_n^1(s) = \sum_{i=1}^{n_1^*(s)} m_i$ and $M_n^0(s) = \sum_{i=1}^{n^*(s)-n_1^*(s)} m_i'$, where $\{m_i\}_{i=1}^{n_1^*(s)}$ and $\{m_i'\}_{i=1}^{n^*(s)-n_1^*(s)}$ are $n_1^*(s)$ i.i.d. multinomial$(1, n_1^{-1}(s), \cdots, n_1^{-1}(s))$ random vectors and $n^*(s) - n_1^*(s)$ i.i.d. multinomial $(1, (n(s) - n_1(s))^{-1}, \cdots, (n(s) - n_1(s))^{-1})$ random vectors, respectively;
2. $M_n^0(s) \perp\!\!\!\perp M_n^1(s)|\mathbb{N}_n$; and
3. $\{M_n^0(s), M_n^1(s)\}_{s \in \mathcal{S}}$ are independent across s given $\mathbb{N}_n$ and are independent of Ψ_n.

Recall that, by Bugni et al. (2018), the original observations can be rearranged according to $s \in \mathcal{S}$ and then within strata, treatment group first and then the control group. Then, given $\mathbb{N}_n$, Step 3 in Section 1.5 implies that the bootstrap observations $\{Y_i^*\}_{i=1}^n$ can be generated by drawing with replacement from the empirical distribution of the outcomes in each (s, a) cell for $(s, a) \in \mathcal{S} \times \{0, 1\}$, $n_a^*(s)$ times, $a = 0, 1$, where $n_0^*(s) = n^*(s) - n_1^*(s)$. Therefore,

$$\sum_{i=1}^{n} A_i^* 1\{S_i^* = s\}\eta_{i,1}^*(s, \tau) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\tilde{\eta}_{i,1}(s, \tau). \tag{1.10.16}$$

Following the standard approach in dealing with the nonparametric bootstrap, we want to approximate

$$M_{ni}, i = N(s) + 1, \cdots, N(s) + n_1(s)$$

by a sequence of i.i.d. Poisson(1) random variables. We construct this sequence as follows. Let $\widetilde{M}_n^1(s) = \sum_{i=1}^{\widetilde{N}(n_1(s))} m_i$, where $\widetilde{N}(k)$ is a Poisson number with mean k and is independent of $\mathbb{N}_n$. The $n_1(s)$ elements of vector $\widetilde{M}_n^1(s)$ is denoted as $\{\widetilde{M}_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$, which is a sequence of i.i.d. Poisson(1) random variables, given $\mathbb{N}_n$. Therefore,

$$\{\widetilde{M}_{ni}, i = N(s) + 1, \cdots, N(s) + n_1(s)|\mathbb{N}_n\} \equiv \{\xi_i^s, i = N(s) + 1, \cdots, N(s) + n_1(s)|\mathbb{N}_n\}$$

where $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$ are i.i.d. sequences of Poisson(1) random variables such that $\{\xi_i^s\}_{i=1}^n$ are independent across $s \in \mathcal{S}$ and against $\mathbb{N}_n$.

Following the argument in van der Vaart and Wellner (1996, Section 3.6), given $n_1(s)$, $n_1^*(s)$, and $\widetilde{N}(n_1(s)) = k$, $|\xi_i^s - M_{ni}|$ is binomially $(|k - n_1^*(s)|, n_1(s)^{-1})$-distributed. In addition, there exists a sequence $\ell_n = O(\sqrt{n(s)})$ such that

$$\begin{aligned}\mathbb{P}(|\widetilde{N}(n_1(s)) - n_1^*(s)| \geq \ell_n) \leq & \mathbb{P}(|\widetilde{N}(n_1(s)) - n_1(s)| \geq \ell_n/3) + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\ \leq & \mathbb{E}\mathbb{P}(|N(n_1(s)) - n_1(s)| \geq \ell_n/3 | n_1(s)) + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\ \leq & \varepsilon/3 + \mathbb{P}(|n_1^*(s) - n_1(s)| \geq 2\ell_n/3) \\ \leq & \varepsilon/3 + \mathbb{P}(|D_n^*(s)| + |D_n(s)| + \pi|n^*(s) - n(s)| \geq 2\ell_n/3) \\ \leq & 2\varepsilon/3 + \mathbb{P}(\pi|n^*(s) - n(s)| \geq \ell_n/3) \\ \leq & \varepsilon,\end{aligned}$$

where the first inequality holds due to the union bound inequality, the second inequality holds by the law of iterated expectation, the third inequality holds because (1) conditionally on data, $\widetilde{N}(n_1(s)) - n_1(s) = O_p(\sqrt{n_1(s)})$ and (2) $n_1(s)/n(s) = \pi + \frac{D_n(s)}{n(s)} \to \pi > 0$ as $n(s) \to \infty$, the fourth inequality holds by the fact that

$$n_1^*(s) - n_1(s) = D_n^*(s) - D_n(s) + \pi(n^*(s) - n(s)),$$

the fifth inequality holds because by Assumptions 1 and 4, $|D_n^*(s)| + |D_n(s)| = O_p(\sqrt{n(s)})$, and the sixth inequality holds because $\{S_i^*\}_{i=1}^n$ is generated from $\{S_i\}_{i=1}^n$ by the standard bootstrap procedure, and thus, by van der Vaart and Wellner (1996, Theorem 3.6.1),

$$n^*(s) - n(s) = \sum_{i=1}^{n}(M_{ni}^w - 1)(1\{S_i = s\} - p(s)) = O_p(\sqrt{n(s)}),$$

where $(M_{n1}^w, \cdots, M_{nn}^w)$ is independent of $\{S_i\}_{i=1}^n$ and multinomially distributed with parameters n and (probabilities) $1/n, \cdots, 1/n$. Therefore, by direct calculation, as $n \to \infty$,

$$\begin{aligned}& \mathbb{P}(\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| > 2) \\ \leq & \mathbb{P}(\max_{N(s)+1 \leq i \leq N(s)+n_1(s)} |\xi_i^s - M_{ni}| > 2, n_1(s) \geq n(s)\varepsilon) + \mathbb{P}(n_1(s) \leq n(s)\varepsilon) \\ \leq & \varepsilon + \mathbb{E} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \mathbb{P}(|\xi_i^s - M_{ni}| > 2, |N(n_1(s)) - n_1^*(s)| \leq \ell_n, n_1(s) \geq n(s)\varepsilon | n_1(s), n_1^*(s), n(s)) + \varepsilon \\ \leq & 2\varepsilon + \mathbb{E} n_1(s)\mathbb{P}(\text{bin}(\ell_n, n_1^{-1}(s)) > 2 | n_1(s), n_1^*(s), n(s)) 1\{n_1(s) \geq n(s)\varepsilon\} \to 2\varepsilon,\end{aligned}$$

where we use the fact that

$$n_1(s)\mathbb{P}(\text{bin}(\ell_n, n_1^{-1}(s)) > 2 | n_1(s), n_1^*(s), n(s)) 1\{n_1(s) \geq n(s)\varepsilon\}$$

$$\lesssim n_1(s)\left(\frac{\ell_n}{n(s)}\right)^3\left(\frac{n(s)}{n_1(s)}\right)^3 1\{n_1(s)\geq n(s)\varepsilon\}\lesssim\frac{1}{\sqrt{n(s)}\varepsilon^3}\to 0.$$

Because ε is arbitrary, we have

$$\mathbb{P}\left(\max_{N(s)+1\leq i\leq N(s)+n_1(s)}|\xi_i^s-M_{ni}|>2\right)\to 0. \tag{1.10.17}$$

Note that $|\xi_i^s-M_{ni}|=\sum_{j=1}^{\infty}1\{|\xi_i^s-M_{ni}|\geq j\}$. Let $I_n^j(s)$ be the set of indexes $i\in\{N(s)+1,\cdots,N(s)+n_1(s)\}$ such that $|\xi_i^s-M_{ni}|\geq j$. Then, $\xi_i^s-M_{ni}=\text{sign}(\widetilde{N}(n_1(s))-n_1^*(s))\sum_{j=1}^{\infty}1\{i\in I_n^j(s)\}$. Thus,

$$\frac{1}{\sqrt{n(s)}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s-M_{ni})\tilde{\eta}_{i,1}(s,\tau)=\text{sign}(\widetilde{N}(n_1(s))-n_1^*(s))\sum_{j=1}^{\infty}\left[\frac{\#I_n^j(s)}{\sqrt{n(s)}}\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\tilde{\eta}_{i,1}(s,\tau)\right]. \tag{1.10.18}$$

In the following, we aim to show that the RHS of (1.10.18) converges to zero in probability uniformly over $s\in\mathcal{S},\tau\in\Upsilon$. First, note that, by (1.10.17), $\max_{N(s)+1\leq i\leq N(s)+n_1(s)}|\xi_i^s-M_{ni}|\leq 2$ occurs with probability approaching one. In the event set that $\max_{N(s)+1\leq i\leq N(s)+n_1(s)}|\xi_i^s-M_{ni}|\leq 2$, only the first two terms of the first summation on the RHS of (1.10.18) can be nonzero. In addition, for any j, we have $j(\#I_n^j(s))\leq|\widetilde{N}(n_1(s))-n_1(s)|=O_p(\sqrt{n(s)})$, and thus, $\frac{\#I_n^j(s)}{\sqrt{n(s)}}=O_p(1)$ for $j=1,2$. Therefore, it suffices to show that, for $j=1,2$,

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}\left|\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\tilde{\eta}_{i,1}(s,\tau)\right|=o_p(1).$$

Note that

$$\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\tilde{\eta}_{i,1}(s,\tau)=\sum_{i=N(s)+1}^{N(s)+n_1(s)}\omega_{ni}\tilde{\eta}_{i,1}(s,\tau), \tag{1.10.19}$$

where $\omega_{ni}=\frac{1\{|\xi_i^s-M_{ni}|\geq j\}}{\#I_n^j(s)}$, $i=N(s)+1,\cdots,N(s)+n_1(s)$ and by construction, $\{\omega_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$ is independent of $\{\eta_{i,1}(s,\tau)\}_{i=1}^{n}$. In addition, because $\{\omega_{ni}\}_{i=N(s)+1}^{N(s)+n_1(s)}$ is exchangeable conditional on $\mathbb{N}_n$, so be it unconditionally. Third, $\sum_{i=N(s)+1}^{N(s)+n_1(s)}\omega_{ni}=1$ and $\max_{i=N(s)+1,\cdots,N(s)+n_1(s)}|\omega_{ni}|\leq 1/\#I_n^j(s)\xrightarrow{p}0$. Then, by the same argument in the proof of van der Vaart and Wellner (1996, Lemma 3.6.16), for some $r \in (0,1)$ and any $n_0 = N(s)+1, \cdots, N(s)+n_1(s)$, we have

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)} \omega_{ni}\tilde{\eta}_{i,1}(s,\tau)\right|^r |\Psi_n, \mathbb{N}_n\right)\\
\leq&(n_0-1)\mathbb{E}\left[\max_{N(s)+n_0\leq i\leq N(s)+n_1(s)} \omega_{ni}^r|\mathbb{N}_n\right]\left[\frac{1}{n_1(s)}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\sup_{\tau\in\Upsilon, s\in\mathcal{S}}|\tilde{\eta}_{i,1}^r(s,\tau)|\right]\\
&+(n_1(s)\mathbb{E}(\omega_{ni}|\mathbb{N}_n))^r \max_{n_0\leq k\leq n_1(s)}\mathbb{E}\left[\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r |\mathbb{N}_n,\Psi_n\right],
\end{aligned}
\tag{1.10.20}
$$

where $(R_{k_1+1}(k_1,k_2), \cdots, R_{k_1+k_2}(k_1,k_2))$ is uniformly distributed on the set of all permutations of $k_1+1, \cdots, k_1+k_2$ and independent of $\mathbb{N}_n$ and Ψ_n. First note that $\sup_{s\in\mathcal{S},\tau\in\Upsilon}|\eta_{i,1}(s,\tau)|$ is bounded and

$$
\max_{N(s)+1\leq i\leq N(s)+n_1(s)} \omega_{ni}^r \leq 1/(\#I_n^j(s))^r \xrightarrow{p} 0.
$$

Therefore, the first term on the RHS of (1.10.20) converges to zero in probability for every fixed n_0. For the second term, because $\omega_{ni}|\mathbb{N}_n$ is exchangeable,

$$
n_1(s)\mathbb{E}(\omega_{ni}|\mathbb{N}_n) = \sum_{i=N(s)+1}^{N(s)+n_1(s)} \mathbb{E}(\omega_{ni}|\mathbb{N}_n) = 1.
$$

In addition, let $\mathbb{S}_n(k_1,k_2)$ be the σ-field generated by all functions of $\{\tilde{\eta}_{i,1}(s,\tau)\}_{i\geq 1}$ that are symmetric in their k_1+1 to k_1+k_2 arguments. Then,

$$
\begin{aligned}
&\max_{n_0\leq k\leq n_1(s)}\mathbb{E}\left[\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r |\mathbb{N}_n,\Psi_n\right]\\
=&\max_{n_0\leq k\leq n_1(s)}\mathbb{E}\left[\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{j,1}(s,\tau)\right|^r |\mathbb{N}_n,\mathbb{S}_n(N(s),n_1(s))\right]\\
\leq&2\mathbb{E}\left\{\max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+1}^{N(s)+k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right] |\mathbb{N}_n,\mathbb{S}_n(N(s),n_1(s))\right\}\\
=&2\mathbb{E}\left\{\max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon, s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right] |\mathbb{N}_n,\mathbb{S}_n(0,n_1(s))\right\},
\end{aligned}
$$

where the inequality holds by the Jansen's inequality and the triangle inequality and the last equality holds because $\{\tilde{\eta}_{j,1}(s,\tau)\}_{j\geq 1}$ is an i.i.d. sequence. Apply expectation on both sides, we obtain that

$$
\begin{aligned}
&\mathbb{E} \max_{n_0\leq k\leq n_1(s)} \mathbb{E}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r |\mathbb{N}_n,\Psi_n\right]\\
\leq & 2\mathbb{E}\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right]. \qquad (1.10.21)
\end{aligned}
$$

By the usual maximal inequality, as $k\to\infty$,

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right| \xrightarrow{a.s.} 0,$$

which implies that as $n_0\to\infty$

$$\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right] \leq \max_{n_0\leq k}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right] \xrightarrow{a.s.} 0.$$

In addition, $\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|$ is bounded. Then, by the bounded convergence theorem, we have, as $n_0\to\infty$,

$$\mathbb{E}\max_{n_0\leq k\leq n}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|^r\right] \to 0.$$

which implies that,

$$\mathbb{E} \max_{n_0\leq k\leq n_1(s)} \mathbb{E}\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=N(s)+n_0}^{N(s)+k}\tilde{\eta}_{R_j(N(s),n_1(s)),1}(s,\tau)\right|^r |\mathbb{N}_n,\Psi_n\right] \xrightarrow{p} 0.$$

Therefore, the second term on the RHS of (1.10.20) converges to zero in probability as $n_0\to\infty$. Then, as $n\to\infty$ followed by $n_0\to\infty$,

$$\mathbb{E}\left(\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)}\omega_{ni}\tilde{\eta}_{i,1}(s,\tau)\right|^r |\Psi_n,\mathbb{N}_n\right) \xrightarrow{p} 0.$$

Hence, by the Markov inequality and (1.10.19), we have

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}\left|\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\tilde{\eta}_{i,1}(s,\tau)\right| \xrightarrow{p} 0.$$

Consequently, following (1.10.18)

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s - M_{ni})\tilde{\eta}_{i,1}(s,\tau)\right| = o_p(\sqrt{n(s)}). \tag{1.10.22}$$

This concludes the first part of this Lemma. For the second part, we note

$$\sum_{i=N(s)+1}^{N(s)+n_1(s)}\widetilde{M}_{ni}\tilde{\eta}_{i,1}(s,\tau) \stackrel{d}{=} \sum_{i=N(s)+1}^{N(s)+n_1(s)}\xi_i^s\tilde{\eta}_{i,1}(s,\tau) \stackrel{d}{=} \sum_{i=1}^{n_1(s)}\xi_i^s\tilde{\eta}_{i,1}(s,\tau),$$

where the second equality holds because $\{\xi_i^s,\tilde{\eta}_{i,1}(s,\tau)\}_{i\geq 1} \perp\!\!\!\perp \{N(s),n_1(s),n(s)\}$. Then, conditionally on $\{N(s),n_1(s),n(s)\}$ and uniformly over $s\in\mathcal{S}$, the usual maximal inequality (van der Vaart and Wellner (1996, Theorem 2.14.1)) implies

$$\sup_{\tau\in\Upsilon}|\sum_{i=N(s)+1}^{N(s)+n_1(s)}\widetilde{M}_{ni}\tilde{\eta}_{i,1}(s,\tau)| \stackrel{d}{=} \sup_{\tau\in\Upsilon}|\sum_{i=1}^{n_1(s)}\xi_i^s\tilde{\eta}_{i,1}(s,\tau)| = O_p(\sqrt{n(s)}). \tag{1.10.23}$$

Combining (1.10.16), (1.10.22), and (1.10.23), we establish (1.10.15). This concludes the proof. □

Lemma 1.10.6. *If Assumptions 1(i) and 1(ii) hold,* $\sup_{s\in\mathcal{S}}\frac{|D_n^*(s)|}{\sqrt{n^*(s)}} = O_p(1)$, $\sup_{s\in\mathcal{S}}\frac{|D_n(s)|}{\sqrt{n(s)}} = O_p(1)$, *and* $n(s)\to\infty$ *for all* $s\in\mathcal{S}$, *a.s., then, uniformly over* $\tau\in\Upsilon$,

$$Q_n^*(u,\tau) \xrightarrow{p} \frac{1}{2}u'Qu.$$

Proof. Recall $Q_{n,1}^*(u,\tau)$ and $Q_{n,0}^*(u,\tau)$ defined in (1.10.10). We focus on $Q_{n,1}^*(u,\tau)$. Recall the definition of M_{ni} in the proof of Lemma 1.10.5. We have

$$Q_{n,1}^*(u,\tau) = \sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}M_{ni}\int_0^{\frac{u_0+u_1}{\sqrt{n}}}(1\{Y_i^s(1)-q_1(\tau)\leq v\} - 1\{Y_i^s(1)-q_1(\tau)\leq 0\})dv$$

$$=\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)]+\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\mathbb{E}\phi_i(u,\tau,s), \tag{1.10.24}$$

where $\phi_i(u,\tau,s)=\int_0^{\frac{u_0+u_1}{\sqrt{n}}}(1\{Y_i^s(1)-q_1(\tau)\leq v\}-1\{Y_i^s(1)-q_1(\tau)\leq 0\})dv$.

Similar to (1.10.22), we have

$$\begin{aligned}&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right]\\ =&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right]+\sum_{s\in\mathcal{S}} r_n(u,\tau,s),\end{aligned} \tag{1.10.25}$$

where $\{\xi_i^s\}_{i=1}^n$ is a sequence of i.i.d. Poisson(1) random variables and is independent of everything else, and

$$r_n(u,\tau,s)=\text{sign}(\widetilde{N}(n_1(s))-n_1^*(s))\sum_{j=1}^{\infty}\frac{\#I_n^j(s)}{\sqrt{n(s)}}\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\sqrt{n(s)}\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right].$$

We aim to show

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|r_n(u,\tau,s)|=o_p(1), \tag{1.10.26}$$

Recall that the proof of Lemma 1.10.5 relies on (1.10.21) and the fact that

$$\mathbb{E}\sup_{n(s)\geq k\geq n_0}\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right|\to 0.$$

Using the same argument and replacing $\tilde{\eta}_{j,1}(s,\tau)$ by $\sqrt{n(s)}\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right]$, in order to show (1.10.26), we only need to verify that, as $n\to\infty$ followed by $n_0\to\infty$,

$$\mathbb{E}\sup_{n(s)\geq k\geq n_0}\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{i=1}^{k}\sqrt{n(s)}\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right]\right|\to 0.$$

Note $\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{i=1}^{k}\sqrt{n(s)}\left[\phi_i(u,\tau,s)-\mathbb{E}\phi_i(u,\tau,s)\right]\right|$ is bounded by $|u_0|+|u_1|$. It suffices to show that, for any $\varepsilon > 0$, as $n(s) \to \infty$ followed by $n_0 \to \infty$,

$$\mathbb{P}\left(\sup_{n(s)\geq k\geq n_0} \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{k}\sum_{i=1}^{k} \sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| \geq \varepsilon \right) \to 0. \tag{1.10.27}$$

Define the class of functions $\mathcal{F}_n$ as

$$\mathcal{F}_n = \{\sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] : \tau \in \Upsilon, s \in \mathcal{S}\}.$$

Then, $\mathcal{F}_n$ is nested by a VC-class with fixed VC-index. In addition, for fixed u, $\mathcal{F}_n$ has a bounded (and independent of n) envelope function $F = |u_0| + |u_1|$. Last, define $\mathcal{I}_l = \{2^l, 2^l+1, \cdots, 2^{l+1}-1\}$. Then,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{n(s)\geq k\geq n_0} \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{k}\sum_{i=1}^{k} \sqrt{n(s)}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right] \right| \geq \varepsilon \middle| n(s) \right) \\
\leq & \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n(s))\rfloor+1} \mathbb{P}\left(\sup_{k\in\mathcal{I}_l} \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{k}\sum_{i=1}^{k} \sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| \geq \varepsilon \middle| n(s) \right) \\
\leq & \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n(s))\rfloor+1} \mathbb{P}\left(\sup_{k\leq 2^{l+1}} \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \sum_{i=1}^{k} \sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| \geq \varepsilon 2^l \middle| n(s) \right) \\
\leq & \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n(s))\rfloor+1} 9\mathbb{P}\left(\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \sum_{i=1}^{2^{l+1}} \sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| \geq \varepsilon 2^l/30 \middle| n(s) \right) \\
\leq & \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n(s))\rfloor+1} \frac{270\mathbb{E}\left(\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \sum_{i=1}^{2^{l+1}} \sqrt{n(s)}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| \middle| n(s) \right)}{\varepsilon 2^l} \\
\leq & \sum_{l=\lfloor \log_2(n_0)\rfloor}^{\lfloor \log_2(n(s))\rfloor+1} \frac{C_1}{\varepsilon 2^{l/2}} \\
\leq & \frac{2C_1}{\varepsilon\sqrt{n_0}} \to 0,
\end{aligned}$$

where the first inequality holds by the union bound, the second inequality holds because on $\mathcal{I}_l$, $2^{l+1} \geq k \geq 2^l$, the third inequality follows the same argument in the proof of Theorem 1.3.1, the fourth inequality is due to the Markov inequality, the fifth inequality follows the standard maximal inequality such as van der Vaart and Wellner (1996, Theorem 2.14.1) and the constant C_1 is independent of (l, ε, n), and the last inequality holds by letting $n \to \infty$. Because ε is arbitrary, we have established (1.10.27), and thus, (1.10.26), which further implies that

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} |r_n(u,\tau,s)| = o_p(1).$$

In addition, for the leading term of (1.10.25), we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s \left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \\
&= \sum_{s\in\mathcal{S}} \left[\Gamma_n^{s*}(N(s)+n_1(s),\tau) - \Gamma_n^{s*}(N(s),\tau)\right],
\end{aligned}$$

where

$$\begin{aligned}
\Gamma_n^{s*}(k,\tau,e) = &\sum_{i=1}^{k} \xi_i^s \int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i^s(1) \le q_1(\tau)+v\} - 1\{Y_i^s(1) \le q_1(\tau)\}\right) dv \\
&- k\mathbb{E}\left[\int_0^{\frac{u_0+u_1}{\sqrt{n}}} \left(1\{Y_i^s(1) \le q_1(\tau)+v\} - 1\{Y_i^s(1) \le q_1(\tau)\}\right) dv\right].
\end{aligned}$$

By the same argument in (1.10.1), we can show that

$$\sup_{0<t\le 1, \tau\in\Upsilon} |\Gamma_n^{s*}(k,\tau,e)| = o_p(1),$$

where we need to use the fact that the Poisson(1) random variable has an exponential tail and thus

$$\mathbb{E} \sup_{i\in\{1,\cdots,n\}, s\in\mathcal{S}} \xi_i^s = O(\log(n)).$$

Therefore,

$$\sup_{\tau\in\Upsilon} \left| \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s) - \mathbb{E}\phi_i(u,\tau,s)\right] \right| = o_p(1). \tag{1.10.28}$$

For the second term on the RHS of (1.10.24), we have

$$\sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\mathbb{E}\phi_i(u,\tau,s) = \sum_{s\in\mathcal{S}} n_1^*(s)\mathbb{E}\phi_i(u,\tau,s)$$

$$=\sum_{s\in\mathcal{S}}\pi p(s)\frac{f_1(q_1(\tau)|s)}{2}(u_0+u_1)^2+o(1)$$

$$=\frac{\pi f_1(q_1(\tau))(u_0+u_1)^2}{2}+o(1), \tag{1.10.29}$$

where the $o(1)$ term holds uniformly over $\tau\in\Upsilon$, the first equality holds because $\sum_{i=N(s)+1}^{N(s)+n_1(s)}M_{ni}=n_1^*(s)$ and the second equality holds by the same calculation in (1.10.1) and the facts that $n^*(s)/n\xrightarrow{p}p(s)$ and

$$\frac{n_1^*(s)}{n}=\frac{D_n^*(s)+\pi n^*(s)}{n}\xrightarrow{p}\pi p(s).$$

Combining (1.10.24)–(1.10.26), (1.10.28), and (1.10.29), we have

$$Q_{n,1}^*(u,\tau)\xrightarrow{p}\frac{\pi f_1(q_1(\tau))(u_0+u_1)^2}{2},$$

uniformly over $\tau\in\Upsilon$. By the same argument, we can show that, uniformly over $\tau\in\Upsilon$,

$$Q_{n,0}^*(u,\tau)\xrightarrow{p}\frac{(1-\pi)f_0(q_0(\tau))u_0^2}{2}.$$

This concludes the proof. □

1.11 Appendix B

1.11.1 Quantile Regression with Strata Fixed Effects

The strata fixed effects estimator for the ATE is obtained by a linear regression of outcome Y_i on the treatment status A_i, controlling for strata dummies $\{1\{S_i=s\}_{s\in\mathcal{S}}\}$. Bugni et al. (2018) point out that, due to the Frisch-Waugh-Lovell theorem, this estimator is equal to the linear coefficient in the regression of Y_i on $\tilde{A}_i$, in which $\tilde{A}_i$ is the residual of the projection of A_i on the strata dummies. Unlike the expectation, the quantile operator is nonlinear. Therefore, we cannot consistently estimate QTEs by a linear QR of Y_i on A_i and strata dummies. Instead, based on the equivalence relationship, we propose to run the QR of Y_i on $\tilde{A}_i$. Formally, let $\tilde{A}_i=A_i-\hat{\pi}(S_i)$ and $\dot{\tilde{A}}_i=(1,\tilde{A}_i)'$, where $\hat{\pi}(s)=n_1(s)/n(s)$, $n_1(s)=\sum_{i=1}^n A_i1\{S_i=s\}$, and $n(s)=\sum_{i=1}^n 1\{S_i=s\}$. Then, the strata fixed effects (SFE) estimator for the QTE is

$\hat{\beta}_{sfe,1}(\tau)$, where

$$\hat{\beta}_{sfe}(\tau) \equiv \left(\hat{\beta}_{sfe,0}(\tau), \hat{\beta}_{sfe,1}(\tau)\right)' = \underset{b=(b_0,b_1)'\in\Re^2}{\arg\min} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \dot{\tilde{A}}_i' b\right).$$

Theorem 1.11.1. *If Assumptions 1(i)–1(iii) and 2 hold and $p(s) > 0$ for $s \in \mathcal{S}$, then, uniformly over $\tau \in \Upsilon$,*

$$\sqrt{n}\left(\hat{\beta}_{sfe,1}(\tau) - q(\tau)\right) \rightsquigarrow \mathcal{B}_{sfe}(\tau), \text{ as } n \to \infty,$$

where $\mathcal{B}_{sfe}(\cdot)$ is a Gaussian process with covariance kernel $\Sigma_{sfe}(\cdot,\cdot)$. The expression for $\Sigma_{sfe}(\cdot,\cdot)$ can be found in the proof of this theorem.

In particular, the asymptotic variance for $\hat{\beta}_{sfe,1}(\tau)$ is

$$\zeta_Y^2(\pi,\tau) + \zeta_A'^2(\pi,\tau) + \zeta_S^2(\tau),$$

where $\zeta_Y^2(\pi,\tau)$ and $\zeta_S^2(\tau)$ are the same as those defined below Theorem 1.3.1,

$$\begin{aligned}\zeta_A'^2(\pi,\tau) =& \mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau) - m_0(S,\tau))\left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right) \\ &+ q(\tau)\left(\frac{f_1(q_1(\tau)|S)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|S)}{f_0(q_0(\tau))}\right)\Bigg]^2.\end{aligned}$$

Three remarks are in order. First, if the treatment assignment rule achieves strong balance, then $\zeta_A'^2(\pi,\tau) = 0$ and the asymptotic variances for $\hat{\beta}_1(\tau)$ and $\hat{\beta}_{sfe,1}(\tau)$ are the same. Second, if the treatment assignment rule does not achieve strong balance, then it is difficult to compare the asymptotic variances of $\hat{\beta}_1(\tau)$ and $\hat{\beta}_{sfe,1}(\tau)$. Based on our simulation results in Section 1.11.6, the SFE estimator usually has a smaller standard error. Third, in order to analytically compute the asymptotic variance $\hat{\beta}_{sfe,1}(\tau)$, one needs to nonparametrically estimate not only the unconditional densities $f_j(\cdot)$ but also the conditional densities $f_j(\cdot|s)$ for $j = 0, 1$ and $s \in \mathcal{S}$. However, such difficulty can be avoided by the covariate-adaptive bootstrap inference considered in Section 1.5.

We can compute the weighted bootstrap counterpart of strata fixed effects estimator:

$$\hat{\beta}_{sfe}^w(\tau) = \arg\min_b \sum_{i=1}^{n} \xi_i \rho_\tau\left(Y_i - \dot{\tilde{A}}_i^{w\prime} b\right),$$

where $\dot{\tilde{A}}_i^w = (1, \tilde{A}_i^w)'$, $\tilde{A}_i^w = A_i - \hat{\pi}^w(S_i)$, and $\hat{\pi}^w(\cdot)$ is defined in Section 1.4. The second element of $\hat{\beta}^w_{sfe}(\tau)$ is our bootstrap estimator of the QTE.

Theorem 1.11.2. *If Assumptions 1–3 hold and $p(s) > 0$ for all $s \in \mathcal{S}$, then uniformly over $\tau \in \Upsilon$ and conditionally on data,*

$$\sqrt{n}\left(\hat{\beta}^w_{sfe,1}(\tau) - \hat{\beta}_{sfe,1}(\tau)\right) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau), \text{ as } n \to \infty,$$

where $\tilde{\mathcal{B}}_{sfe}(\tau)$ is a Gaussian process with covariance kernel being equal to that of $\mathcal{B}_{sfe}(\tau)$ defined in Theorem 1.11.1 with $\gamma(s)$ being replaced by $\pi(1-\pi)$.

Similar to the SQR estimator, the weighted bootstrap fails to capture the cross-sectional dependence due to the covariate-adaptive randomization, and thus, overestimates the asymptotic variance of the SFE estimator.

We can also implement the covariate-adaptive bootstrap. Let

$$\hat{\beta}^*_{sfe}(\tau) = \arg\min_b \sum_{i=1}^n \rho_\tau\left(Y_i^* - \dot{\tilde{A}}_i^{*\prime} b\right),$$

where $\dot{\tilde{A}}_i^* = (1, \tilde{A}_i^*)'$, $\tilde{A}_i^* = A_i^* - \hat{\pi}^*(S_i^*)$, $\hat{\pi}^*(s) = \frac{n_1^*(s)}{n^*(s)}$, and $(Y_i^*, A_i^*, S_i^*)_{i=1}^n$ is the covariate-adaptive bootstrap sample generated via the procedure mentioned in Section 1.5. The the second element $\hat{\beta}^*_{sfe,1}(\tau)$ of $\hat{\beta}^*_{sfe}(\tau)$ is the covariate-adaptive SFE estimator.

Theorem 1.11.3. *If Assumptions 1, 2, and 4 hold and $p(s) > 0$ for all $s \in \mathcal{S}$, then, uniformly over $\tau \in \Upsilon$ and conditionally on data,*

$$\sqrt{n}\left(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau)\right) \rightsquigarrow \mathcal{B}_{sfe}(\tau), \text{ as } n \to \infty.$$

Unlike the weighted bootstrap, the covariate-adaptive bootstrap can mimic the cross-sectional dependence, and thus, produces an asymptotically valid standard error for the SFE estimator.

1.11.2 Proof of Theorem 1.11.1

Define $\tilde{\beta}_1(\tau) = q(\tau)$, $\tilde{\beta}_0(\tau) = \pi q_1(\tau) + (1-\pi) q_0(\tau)$, $\tilde{\beta}(\tau) = (\tilde{\beta}_0(\tau), \tilde{\beta}_1(\tau))'$, and $\breve{A}_i = (1, A_i - \pi)'$. For arbitrary b_0 and b_1, let $u_0 = \sqrt{n}(b_0 - \tilde{\beta}_0(\tau))$, $u_1 = \sqrt{n}(b_1 - \tilde{\beta}_1(\tau))$, $u = (u_0, u_1)' \in \Re^2$, and

$$L_{sfe,n}(u,\tau) = \sum_{i=1}^n \left[\rho_\tau(Y_i - \breve{A}_i'\tilde{\beta}(\tau) - (\dot{\tilde{A}}_i' b - \breve{A}_i'\tilde{\beta}(\tau))) - \rho_\tau(Y_i - \breve{A}_i'\tilde{\beta}(\tau))\right].$$

Then, by the change of variable, we have that

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) = \arg\min_u L_{sfe,n}(u, \tau).$$

Notice that $L_{sfe,n}(u,\tau)$ is convex in u for each τ and bounded in τ for each u. In the following, we aim to show that there exists

$$g_{sfe,n}(u,\tau) = -u'W_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}(\tau)u$$

such that (1) for each u,

$$\sup_{\tau\in\Upsilon} |L_{sfe,n}(u,\tau) - g_{sfe,n}(u,\tau) - h_{sfe,n}(\tau)| \xrightarrow{p} 0,$$

where $h_{sfe,n}(\tau)$ does not depend on u; (2) the maximum eigenvalue of $Q_{sfe}(\tau)$ is bounded from above and the minimum eigenvalue of $Q_{sfe}(\tau)$ is bounded away from 0 uniformly over $\tau \in \Upsilon$; (3) $W_{sfe,n}(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau)$ uniformly over $\tau \in \Upsilon$ for some $\tilde{\mathcal{B}}(\tau)$.[10] Then by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1}W_{sfe,n}(\tau) + r_{sfe,n}(\tau),$$

where $\sup_{\tau\in\Upsilon} ||r_{sfe,n}(\tau)|| = o_p(1)$. In addition, by (3), we have, uniformly over $\tau \in \Upsilon$,

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) \rightsquigarrow [Q_{sfe}(\tau)]^{-1}\tilde{\mathcal{B}}(\tau) \equiv \mathcal{B}(\tau).$$

The second element of $\mathcal{B}(\tau)$ is $\mathcal{B}_{sfe}(\tau)$ stated in Theorem 1.11.1. Next, we prove requirements (1)–(3) in three steps.

Step 1. By Knight's identity (Knight, 1998), we have

$$\begin{aligned}
&L_{sfe,n}(u,\tau)\\
=& - \sum_{i=1}^{n} (\dot{\breve{A}}_i'(\tilde{\beta}(\tau) + \frac{u}{\sqrt{n}}) - \breve{A}_i'\tilde{\beta}(\tau)) \left(\tau - 1\{Y_i \le \dot{\breve{A}}_i'\tilde{\beta}(\tau)\}\right)\\
&+ \sum_{i=1}^{n} \int_0^{\dot{\breve{A}}_i'(\tilde{\beta}(\tau)+\frac{u}{\sqrt{n}})-\breve{A}_i'\tilde{\beta}(\tau)} \left(1\{Y_i - \dot{\breve{A}}_i'\tilde{\beta}(\tau) \le v\} - 1\{Y_i - \dot{\breve{A}}_i'\tilde{\beta}(\tau) \le 0\}\right) dv\\
\equiv& - L_{1,n}(u,\tau) + L_{2,n}(u,\tau).
\end{aligned}$$

[10] We abuse the notation and denote the weak limit of $W_{sfe,n}(\tau)$ as $\tilde{\mathcal{B}}(\tau)$. This limit is different from the weak limit of $W_n(\tau)$ in the proof of Theorem 1.3.1.

Step 1.1. We first consider $L_{1,n}(u,\tau)$. Note that $\tilde{\beta}_1(\tau) = q(\tau)$ and

$$\begin{aligned}
&L_{1,n}(u,\tau)\\
=&\sum_{i=1}^n\sum_{s\in\mathcal{S}} A_i 1\{S_i=s\}\left(\frac{u_0}{\sqrt{n}} + (1-\hat{\pi}(s))\frac{u_1}{\sqrt{n}} + (\pi-\hat{\pi}(s))q(\tau)\right)(\tau - 1\{Y_i(1)\le q_1(\tau)\})\\
&+\sum_{i=1}^n\sum_{s\in\mathcal{S}}(1-A_i)1\{S_i=s\}\left(\frac{u_0}{\sqrt{n}} - \hat{\pi}(s)\frac{u_1}{\sqrt{n}} + (\pi-\hat{\pi}(s))q(\tau)\right)(\tau - 1\{Y_i(0)\le q_0(\tau)\})\\
\equiv&L_{1,1,n}(u,\tau) + L_{1,0,n}(u,\tau). \qquad (1.11.1)
\end{aligned}$$

Let $\iota_1 = (1, 1-\pi)'$ and $\iota_0 = (1,-\pi)'$. Note that $\hat{\pi}(s) - \pi = \frac{D_n(s)}{n(s)}$. Then, for $L_{1,1,n}(u,\tau)$, we have

$$\begin{aligned}
&L_{1,1,n}(u,\tau)\\
=&\sum_{i=1}^n\sum_{s\in\mathcal{S}} A_i 1\{S_i=s\}\left[\frac{u'\iota_1}{\sqrt{n}} + (\pi-\hat{\pi}(s))\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)\right](\tau - 1\{Y_i(1)\le q_1(\tau)\})\\
=&\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\sum_{s\in\mathcal{S}} A_i 1\{S_i=s\}(\tau - 1\{Y_i(1)\le q_1(\tau)\})\\
&-\sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}}\frac{u_1}{n(s)}\sum_{i=1}^n A_i 1\{S_i=s\}(\tau - 1\{Y_i(1)\le q_1(\tau)\})\\
&+\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}(s))q(\tau)\sum_{i=1}^n A_i 1\{S_i=s\}(\tau - 1\{Y_i(1)\le q_1(\tau)\})\\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau) + (A_i-\pi)1\{S_i=s\}m_1(s,\tau) + \pi 1\{S_i=s\}m_1(s,\tau)\Big]\\
&-\sum_{s\in\mathcal{S}}\frac{D_n(s)}{\sqrt{n}}\frac{u_1}{n(s)}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau) + (A_i-\pi)1\{S_i=s\}m_1(s,\tau) + \pi 1\{S_i=s\}m_1(s,\tau)\Big] + h_{1,}\\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\Big[A_i1\{S_i=s\}\eta_{i,1}(s,\tau) + (A_i-\pi)1\{S_i=s\}m_1(s,\tau) + \pi 1\{S_i=s\}m_1(s,\tau)\Big]\\
&-\sum_{s\in\mathcal{S}}\frac{u_1D_n(s)\pi m_1(s,\tau)}{\sqrt{n}} + h_{1,1}(\tau) + R_{sfe,1,1}(u,\tau), \qquad (1.11.2)
\end{aligned}$$

where

$$h_{1,1}(\tau) = \sum_{s\in\mathcal{S}}(\pi-\hat{\pi}(s))q(\tau)\sum_{i=1}^n A_i1\{S_i=s\}(\tau - 1\{Y_i(1)\le q_1(\tau)\})$$

and

$$R_{sfe,1,1}(u,\tau) = -\sum_{s\in\mathcal{S}} \frac{u_1 D_n(s)}{\sqrt{n}n(s)} \sum_{i=1}^{n} \Big[A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + (A_i - \pi)1\{S_i = s\}m_1(s,\tau)\Big].$$

By the same argument in Lemma 1.10.2 and Assumption 1(iii), we have for every $s \in \mathcal{S}$,

$$\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)\right| = O_p(1) \tag{1.11.3}$$

and

$$\sup_{\tau\in\Upsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[(A_i - \pi)1\{S_i = s\}m_1(s,\tau)\Big]\right| = \sup_{\tau\in\Upsilon}\left|\frac{D_n(s)m_1(s,\tau)}{\sqrt{n}}\right| = O_p(1).$$

In addition, note that $n(s)/n \xrightarrow{p} p(s)$. Therefore,

$$\sup_{\tau\in\Upsilon}|R_{sfe,1,1}(u,\tau)| = O_p(\frac{1}{\sqrt{n}}) = o_p(1).$$

Similarly, we have

$$\begin{aligned}
&L_{1,0,n}(u,\tau) \\
=&\sum_{s\in\mathcal{S}} \frac{u'\iota_0}{\sqrt{n}} \sum_{i=1}^{n}\Big[(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau) - (A_i - \pi)1\{S_i = s\}m_0(s,\tau) + (1-\pi)1\{S_i = s\}m_0(s,\tau)\Big] \\
&-\sum_{s\in\mathcal{S}} \frac{u_1 D_n(s)(1-\pi)m_0(s,\tau)}{\sqrt{n}} + h_{1,0}(\tau) + R_{sfe,1,0}(u,\tau),
\end{aligned} \tag{1.11.4}$$

where

$$h_{1,0}(\tau) = \sum_{s\in\mathcal{S}}(\pi - \hat{\pi}(s))q(\tau)\sum_{i=1}^{n}(1-A_i)1\{S_i = s\}\left(\tau - 1\{Y_i(0) \leq q_0(\tau)\}\right),$$

$$R_{sfe,1,0}(u,\tau) = -\sum_{s\in\mathcal{S}} \frac{u_1 D_n(s)}{\sqrt{n}n(s)} \sum_{i=1}^{n}\Big[(1-A_i)1\{S_i = s\}\eta_{i,0}(\tau) - (A_i - \pi)1\{S_i = s\}m_0(s,\tau)\Big],$$

and

$$\sup_{\tau \in \Upsilon} |R_{sfe,1,0}(\tau)| = O_p(\frac{1}{\sqrt{n}}) = o_p(1).$$

Combining (1.11.1), (1.11.2), (1.11.4) and letting $\iota_2 = (1, 1-2\pi)'$, we have

$$\begin{aligned}
L_{1,n}(u,\tau) =& \frac{1}{\sqrt{n}} \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \Big[u'\iota_1 A_i 1\{S_i = s\} \eta_{i,1}(s,\tau) + u'\iota_0 (1-A_i) 1\{S_i = s\} \eta_{i,0}(s,\tau) \Big] \\
&+ \sum_{s \in \mathcal{S}} u'\iota_2 \frac{D_n(s)}{\sqrt{n}} \left(m_1(s,\tau) - m_0(s,\tau) \right) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(u'\iota_1 \pi m_1(S_i,\tau) + u'\iota_0 (1-\pi) m_0(S_i,\tau) \right) \\
&+ R_{sfe,1,1}(u,\tau) + R_{sfe,1,0}(u,\tau) + h_{1,1}(\tau) + h_{1,0}(\tau).
\end{aligned} \tag{1.11.5}$$

Step 1.2. Next, we consider $L_{2,n}(u,\tau)$. Denote $E_n(s) = \sqrt{n}(\hat{\pi}(s) - \pi)$. Then,

$$\{E_n(s)\}_{s \in \mathcal{S}} = \left\{ \frac{D_n(s)}{\sqrt{n}} \frac{n}{n(s)} \right\}_{s \in \mathcal{S}} \rightsquigarrow \mathcal{N}(0, \Sigma_D') = O_p(1),$$

where $\Sigma_D' = \text{diag}(\gamma(s)/p(s) : s \in \mathcal{S})$. In addition,

$$\begin{aligned}
& L_{2,n}(u,\tau) \\
=& \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} A_i 1\{S_i = s\} \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i(1) \le q_1(\tau) + v\} - 1\{Y_i(1) \le q_1(\tau)\} \right) dv \\
&+ \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} (1-A_i) 1\{S_i = s\} \int_0^{\frac{u'\iota_0}{\sqrt{n}} - \frac{E_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i(0) \le q_0(\tau) + v\} - 1\{Y_i(0) \le q_0(\tau)\} \right) dv \\
\equiv& L_{2,1,n}(u,\tau) + L_{2,0,n}(u,\tau).
\end{aligned} \tag{1.11.6}$$

By the same argument in (1.10.1), we have

$$\begin{aligned}
L_{2,1,n}(u,\tau) \stackrel{d}{=}& \sum_{s \in \mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i^s(1) \le q_1(\tau) + v\} - 1\{Y_i^s(1) \le q_1(\tau)\} \right) dv \\
\equiv& \sum_{s \in \mathcal{S}} \left[\Gamma_n^s(N(s) + n_1(s), \tau, E_n(s)) - \Gamma_n^s(N(s), \tau, E_n(s)) \right],
\end{aligned} \tag{1.11.7}$$

where

$$\Gamma_n^s(k,\tau,e) = \sum_{i=1}^{k} \int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1) \leq q_1(\tau)+v\} - 1\{Y_i^s(1) \leq q_1(\tau)\}\right) dv.$$

We want to show, for some any sufficiently large constant M,

$$\sup_{0<t\leq 1, \tau\in\Upsilon, |e|\leq M} |\Gamma_n^s(\lfloor nt \rfloor,\tau,e) - \mathbb{E}\Gamma_n^s(\lfloor nt \rfloor,\tau,e)| = o_p(1). \tag{1.11.8}$$

By the same argument in (1.10.2), it suffices to show that

$$\sup_{\tau\in\Upsilon,|e|\leq M} n||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} = o_p(1),$$

where

$$\mathcal{F} = \left\{ \int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1) \leq q_1(\tau)+v\} - 1\{Y_i^s(1) \leq q_1(\tau)\}\right) dv : \tau \in \Upsilon, |e| \leq M \right\}$$

with an envelope $F = \frac{|u_0|+|u_1|+M\sup_{\tau\in\Upsilon}|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}}$. Note that

$$\begin{aligned}\sup_{f\in\mathcal{F}} \mathbb{E}f^2 \leq & \sup_{\tau\in\Upsilon} \mathbb{E}\left[\frac{|u_0|+|u_1|+M|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}} 1\left\{|Y_i^s(1) - q_1(\tau)| \leq \frac{|u_0|+|u_1|+M|q(\tau)|+\frac{|u_1|}{\sqrt{n}}}{\sqrt{n}}\right\}\right]^2 \\ \lesssim & n^{-3/2},\end{aligned}$$

and $\mathcal{F}$ is a VC-class with a fixed VC index. Then, by Chernozhukov et al. (2014, Corollary 5.1),

$$\mathbb{E} \sup_{\tau\in\Upsilon,|e|\leq M} |\Gamma_n^s(n,\tau,e) - \mathbb{E}\Gamma_n^s(n,\tau,e)| = n||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}} \lesssim n\left[\sqrt{\frac{\log(n)}{n^{5/2}}} + \frac{\log(n)}{n^{3/2}}\right] = o(1). \tag{1.11.9}$$

In addition, we have

$$\mathbb{E}\Gamma_n^s(\lfloor nt \rfloor,\tau,e) = \lfloor nt \rfloor \int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} [F_1(q_1(\tau)+v|s) - F_1(q_1(\tau)|s)]dv$$

$$=t\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1 - eq(\tau))^2 + o(1), \tag{1.11.10}$$

where $F_j(\cdot|s)$ and $f_j(\cdot|s)$, $j = 0, 1$ are the conditional CDF and PDF for $Y(j)$ given $S = s$, respectively, and the $o(1)$ term holds uniformly over $\{\tau \in \Upsilon, |e| \leq M\}$. Combining (1.11.8) and (1.11.10) with the fact that $\frac{n_1(s)}{n} \xrightarrow{p} \pi p(s)$, we have

$$\begin{aligned} L_{2,1,n}(u,\tau) =& \sum_{s\in\mathcal{S}} \pi p(s) \frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1 - E_n(s)q(\tau))^2 + R'_{sfe,2,1}(u,\tau) \\ =& \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n(s)u'\iota_1}{\sqrt{n}}q(\tau) + h_{2,1}(\tau) + R_{sfe,2,1}(u,\tau), \end{aligned} \tag{1.11.11}$$

where

$$\sup_{\tau\in\Upsilon} |R'_{sfe,2,1}(u,\tau)| = o_p(1), \quad \sup_{\tau\in\Upsilon} |R_{sfe,2,1}(u,\tau)| = o_p(1),$$

and

$$h_{2,1}(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)E_n^2(s)\tilde{\beta}_1^2(\tau).$$

Similarly, we have

$$\begin{aligned} L_{2,0,n}(u,\tau) =& \frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2 - \sum_{s\in\mathcal{S}}(1-\pi)f_0(q_0(\tau)|s)\frac{D_n(s)u'\iota_0}{\sqrt{n}}q(\tau) \\ &+ h_{2,0}(\tau) + R_{sfe,2,0}(u,\tau), \end{aligned} \tag{1.11.12}$$

where

$$\sup_{\tau\in\Upsilon} |R_{sfe,2,0}(u,\tau)| = o_p(1) \quad \text{and} \quad h_{2,0}(\tau) = \sum_{s\in\mathcal{S}} \frac{(1-\pi)f_0(q_0(\tau)|s)}{2} p(s)E_n^2(s)\tilde{\beta}_1^2(\tau).$$

Combining (1.11.6), (1.11.11), and (1.11.12), we have

$$\begin{aligned} L_{2,n}(u,\tau) =& \frac{1}{2}u'Q_{sfe}(\tau)u - \sum_{s\in\mathcal{S}} q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1 + f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D_n(s)}{\sqrt{n}} \\ &+ R_{sfe,2,1}(u,\tau) + R_{sfe,2,0}(u,\tau) + h_{2,1}(\tau) + h_{2,0}(\tau). \end{aligned} \tag{1.11.13}$$

where

$$
\begin{aligned}
Q_{sfe} =&\pi f_1(q_1(\tau))\iota_1\iota_1' + (1-\pi)f_0(q_0(\tau))\iota_0\iota_0' \\
=&\begin{pmatrix} \pi f_1(q_1(\tau)) + (1-\pi)f_0(q_0(\tau)) & \pi(1-\pi)(f_1(q_1(\tau)) - f_0(q_0(\tau))) \\ \pi(1-\pi)(f_1(q_1(\tau)) - f_0(q_0(\tau))) & \pi(1-\pi)((1-\pi)f_1(q_1(\tau)) + \pi f_0(q_0(\tau))) \end{pmatrix}.
\end{aligned}
$$

Step 1.3. Last, by combining (1.11.5) and (1.11.13), we have

$$
L_{sfe,n}(u,\tau) = -u'W_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}(\tau)u + R_{sfe}(u,\tau) + h_{sfe,n}(\tau),
$$

where

$$
\begin{aligned}
&W_{sfe,n}(\tau) \\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\Big[\iota_1 A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) + \iota_0(1-A_i)1\{S_i = s\}\eta_{i,0}(s,\tau)\Big] \\
&+\sum_{s\in\mathcal{S}}\Big\{\iota_2\left(m_1(s,\tau) - m_0(s,\tau)\right) + q(\tau)\Big[f_1(q_1(\tau)|s)\pi\iota_1 + f_0(q_0(\tau)|s)(1-\pi)\iota_0\Big]\Big\}\frac{D_n(s)}{\sqrt{n}} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\iota_1\pi m_1(S_i,\tau) + \iota_0(1-\pi)m_0(S_i,\tau)\right) \\
\equiv&W_{sfe,n,1}(\tau) + W_{sfe,n,2}(\tau) + W_{sfe,n,3}(\tau),
\end{aligned} \tag{1.11.14}
$$

$$
R_{sfe}(u,\tau) = R_{sfe,1,1}(u,\tau) + R_{sfe,1,0}(u,\tau) + R_{sfe,2,1}(u,\tau) + R_{sfe,2,0}(u,\tau)
$$

such that $\sup_{\tau\in\Upsilon}|R_{sfe}(u,\tau)| = o_p(1)$, and

$$
h_{sfe,n}(\tau) = h_{1,1}(\tau) + h_{1,0}(\tau) + h_{2,1}(\tau) + h_{2,0}(\tau).
$$

This concludes the proof of Step 1.

Step 2. Note that $\det(Q_{sfe}(\tau)) = \pi(1-\pi)f_0(q_0(\tau))f_1(q_1(\tau))$, which is bounded and bounded away from zero. In addition, it can be shown that the two eigenvalues of $Q_{sfe}(\tau)$ are nonnegative. This leads to the desired result.

Step 3. Lemma 1.11.1 establishes the weak convergence that

$$
(W_{sfe,1,n}(\tau), W_{sfe,2,n}(\tau), W_{sfe,3,n}(\tau)) \rightsquigarrow (\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau)),
$$

where $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are three independent two-dimensional Gaussian processes with covariance kernels $\Sigma_1(\tau_1,\tau_2)$, $\Sigma_2(\tau_1,\tau_2)$, and $\Sigma_3(\tau_1,\tau_2)$, respectively. Therefore, uniformly over $\tau \in \Upsilon$,

$$W_{sfe,n}(\tau) \rightsquigarrow \tilde{\mathcal{B}}(\tau),$$

where $\tilde{\mathcal{B}}(\tau)$ is a two-dimensional Gaussian process with covariance kernel

$$\tilde{\Sigma}(\tau_1,\tau_2) = \sum_{j=1}^{3} \Sigma_j(\tau_1,\tau_2).$$

Consequently,

$$\sqrt{n}(\hat{\beta}_{sfe}(\tau) - \tilde{\beta}(\tau)) \rightsquigarrow \mathcal{B}(\tau) \equiv Q_{sfe}^{-1}(\tau)\tilde{\mathcal{B}}(\tau),$$

where $\Sigma(\tau_1,\tau_2)$, the covariance kernel of $\mathcal{B}(\tau)$, has the expression that

$$\begin{aligned}
&\Sigma(\tau_1,\tau_2)\\
=&Q_{sfe}^{-1}(\tau_1)\tilde{\Sigma}(\tau_1,\tau_2)Q_{sfe}^{-1}(\tau_2)\\
=&\Bigg\{\frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\left[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)\right]\begin{pmatrix}\pi^2 & \pi\\ \pi & 1\end{pmatrix}\\
&+\frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\left[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)\right]\begin{pmatrix}(1-\pi)^2 & \pi-1\\ \pi-1 & 1\end{pmatrix}\Bigg\}\\
&+\Bigg\{\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_1))}+\frac{1-\pi}{f_1(q_1(\tau_1))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\end{pmatrix}+q(\tau_1)\frac{f_1(q_1(\tau_1)|S)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\1\end{pmatrix}\\
&+q(\tau_1)\frac{f_0(q_0(\tau_1)|S)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\times\Bigg[(m_1(S,\tau_2)-m_0(S,\tau_2))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_2))}+\frac{1-\pi}{f_1(q_1(\tau_2))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\end{pmatrix}\\
&+q(\tau_2)\frac{f_1(q_1(\tau_2)|S)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\1\end{pmatrix}+q(\tau_2)\frac{f_0(q_0(\tau_2)|S)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\Bigg\}\\
&+\Bigg\{\mathbb{E}\Bigg[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\1\end{pmatrix}+\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\Bigg[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\1\end{pmatrix}+\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]'\Bigg\}.
\end{aligned}$$

By checking the $(2,2)$-element of $\Sigma(\tau_1,\tau_2)$, we have

$$\Sigma_{sfe}(\tau_1,\tau_2)$$

$$
\begin{aligned}
=&\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\\
&+\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right)+q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))}-\frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right)\Bigg]\\
&\times\Bigg[(m_1(S,\tau_2)-m_0(S,\tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right)+q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))}-\frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\Bigg]\\
&+\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right].
\end{aligned}
$$

1.11.3 Proof of Theorem 1.11.2

Note that

$$\sqrt{n}(\hat{\beta}^w_{sfe}(\tau)-\tilde{\beta}(\tau))=\arg\min_u L^w_{sfe,n}(u,\tau),$$

where

$$L^w_{sfe,n}(u,\tau)=\sum_{i=1}^n\xi_i\left[\rho_\tau(Y_i-\dot{\tilde{A}}^{w\prime}_i(\tilde{\beta}(\tau)+\frac{u}{\sqrt{n}}))-\rho_\tau(Y_i-\breve{A}_i'\tilde{\beta}(\tau))\right],$$

$\dot{\tilde{A}}^w_i=(1,\tilde{A}^w_i)'$, $\tilde{A}^w_i=A_i-\hat{\pi}^w(S_i)$, and

$$\hat{\pi}^w(s)=\frac{\sum_{i=1}^n\xi_iA_i1\{S_i=s\}}{\sum_{i=1}^n\xi_i1\{S_i=s\}}.$$

Similar to the proof of Theorem 1.11.1, we divide the proof into two steps. In the first step, we show that there exists

$$g^w_{sfe,n}(u,\tau)=-u'W^w_{sfe,n}(\tau)+\frac{1}{2}u'Q_{sfe}(\tau)u$$

and $h^w_{sfe,n}(\tau)$ independent of u such that for each u

$$\sup_{\tau\in\Upsilon}|L^w_{sfe,n}(u,\tau)-g^w_{sfe,n}(u,\tau)-h^w_{sfe,n}(\tau)|\xrightarrow{p}0.$$

In addition, we will show that $\sup_{\tau\in\Upsilon}||W^w_{sfe,n}(\tau)||=O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w_{sfe}(\tau)-\tilde{\beta}(\tau))=[Q_{sfe}(\tau)]^{-1}W^w_{sfe,n}(\tau)+R^w_{sfe,n}(\tau),$$

where

$$\sup_{\tau \in \Upsilon} ||R^w_{sfe,n}(\tau)|| = o_p(1).$$

In the second step, we show that, conditionally on data,

$$\sqrt{n}(\hat{\beta}^w_{sfe,1}(\tau) - \hat{\beta}_{sfe,1}(\tau)) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau).$$

Step 1. Following Step 1 in the proof of Theorem 1.11.1, we have

$$L^w_{sfe,n}(u,\tau) \equiv -L^w_{1,n}(u,\tau) + L^w_{2,n}(u,\tau),$$

where

$$\begin{aligned}
&L^w_{1,n}(u,\tau) \\
=&\sum_{i=1}^n \sum_{s \in \mathcal{S}} \xi_i A_i 1\{S_i = s\} \left(\frac{u_0}{\sqrt{n}} + (1 - \hat{\pi}^w(s)) \frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^w(s)) q(\tau) \right) (\tau - 1\{Y_i \le q_1(\tau)\}) \\
&+ \sum_{i=1}^n \sum_{s \in \mathcal{S}} \xi_i (1 - A_i) 1\{S_i = s\} \left(\frac{u_0}{\sqrt{n}} - \hat{\pi}^w(s) \frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^w(s)) q(\tau) \right) (\tau - 1\{Y_i \le q_0(\tau)\}) \\
\equiv& L^w_{1,1,n}(u,\tau) + L^w_{1,0,n}(u,\tau),
\end{aligned}$$

$$\begin{aligned}
&L^w_{2,n}(u,\tau) \\
=&\sum_{s \in \mathcal{S}} \sum_{i=1}^n \xi_i A_i 1\{S_i = s\} \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} (1\{Y_i \le q_1(\tau) + v\} - 1\{Y_i \le q_1(\tau)\})\, dv \\
&+ \sum_{s \in \mathcal{S}} \sum_{i=1}^n \xi_i (1 - A_i) 1\{S_i = s\} \int_0^{\frac{u'\iota_0}{\sqrt{n}} - \frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} (1\{Y_i \le q_0(\tau) + v\} - 1\{Y_i \le q_0(\tau)\})\, dv \\
\equiv& L^w_{2,1,n}(u,\tau) + L^w_{2,0,n}(u,\tau),
\end{aligned}$$

and $E^w_n(s) = \sqrt{n}(\hat{\pi}^w(s) - \pi)$.

Step 1.1. Recall that $\iota_1 = (1, 1-\pi)'$ and $\iota_0 = (1, -\pi)'$. In addition, denote $\hat{\pi}^w(s) - \pi = \frac{D^w_n(s)}{n^w(s)}$, where

$$D^w_n(s) = \sum_{i=1}^n \xi_i (A_i - \pi) 1\{S_i = s\} \quad \text{and} \quad n^w(s) = \sum_{i=1}^n \xi_i 1\{S_i = s\}.$$

Then, we have

$$
\begin{aligned}
&L^w_{1,1,n}(u,\tau) \\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_1}{\sqrt{n}}\sum_{i=1}^n\xi_i\left[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+\pi1\{S_i=s\}m_1(s,\tau)\right]+\sum_{s\in\mathcal{S}}\frac{u'\iota_2D^w_n(s)m_1(s,\tau)}{\sqrt{n}} \\
&+h^w_{1,1}(\tau)+R^w_{sfe,1,1}(u,\tau), \qquad (1.11.15)
\end{aligned}
$$

where $\eta_{i,1}(s,\tau)=(\tau-1\{Y_i(1)\leq q_1(\tau)\})-m_1(s,\tau)$,

$$
h^w_{1,1}(\tau)=\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}^w(s))q(\tau)\left(\sum_{i=1}^n\xi_iA_i1\{S_i=s\}(\tau-1\{Y_i\leq q_1(\tau)\})\right),
$$

and

$$
R^w_{sfe,1,1}(u,\tau)=-\sum_{s\in\mathcal{S}}\frac{u_1D^w_n(s)}{\sqrt{n}n^w(s)}\left\{\sum_{i=1}^n\xi_i\left[A_i1\{S_i=s\}\eta_{i,1}(s,\tau)+(A_i-\pi)1\{S_i=s\}m_1(s,\tau)\right]\right\}. \qquad (1.11.16)
$$

By Lemma 1.11.2, we have

$$
\sup_{\tau\in\Upsilon}|R^w_{sfe,1,1}(u,\tau)|=o_p(1).
$$

Similarly, we have

$$
\begin{aligned}
&L^w_{1,0,n}(u,\tau) \\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^n\xi_i\left\{\frac{u'\iota_0}{\sqrt{n}}\left[(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)+\pi1\{S_i=s\}m_1(s,\tau)\right]-\frac{u'\iota_2}{\sqrt{n}}(A_i-\pi)1\{S_i=s\}m_0(s,\tau)\right\} \\
&+h^w_{1,0}(\tau)+R^w_{sfe,1,0}(u,\tau), \qquad (1.11.17)
\end{aligned}
$$

where

$$
\sup_{\tau\in\Upsilon}|R^w_{sfe,1,0}(u,\tau)|=o_p(1).
$$

Combining (1.11.15) and (1.11.17), we have

$$
L^w_{1,n}(u,\tau)
$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i\Big[u'\iota_1 A_i 1\{S_i=s\}\eta_{i,1}(u,\tau)+u'\iota_0(1-A_i)1\{S_i=s\}\eta_{i,0}(u,\tau)\\
&+u'\iota_2(A_i-\pi)1\{S_i=s\}(m_1(s,\tau)-m_0(s,\tau))+1\{S_i=s\}(u'\iota_1\pi m_1(s,\tau)+u'\iota_0(1-\pi)m_0(s,\tau))\Big]\\
&+R^w_{sfe,1,1}(u,\tau)+R^w_{sfe,1,0}(u,\tau)+h^w_{1,1}(\tau)+h^w_{1,0}(\tau).
\end{aligned}
$$

Furthermore, by Lemma 1.11.3, we have

$$
L^w_{2,1,n}(u,\tau)=\frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2-\sum_{s\in\mathcal{S}}f_1(q_1(\tau)|s)\frac{\pi D^w_n(s)u'\iota_1}{\sqrt{n}}q(\tau)+h^w_{2,1}(\tau)+R^w_{sfe,2,1}(u,\tau)
\tag{1.11.18}
$$

and

$$
L^w_{2,0,n}(u,\tau)=\frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2-\sum_{s\in\mathcal{S}}f_0(q_0(\tau)|s)\frac{(1-\pi)D^w_n(s)u'\iota_0}{\sqrt{n}}q(\tau)+h^w_{2,0}(\tau)+R^w_{sfe,2,0}(u,\tau),
\tag{1.11.19}
$$

where

$$
h^w_{2,1}(\tau)=\sum_{s\in\mathcal{S}}\frac{\pi f_1(q_1(\tau)|s)}{2}p(s)(E^w_n(s))^2q^2(\tau),
$$

$$
h^w_{2,0}(\tau)=\sum_{s\in\mathcal{S}}\frac{(1-\pi)f_0(q_0(\tau)|s)}{2}p(s)(E^w_n(s))^2q^2(\tau),
$$

$$
\sup_{\tau\in\Upsilon}|R^w_{sfe,2,1}(u,\tau)|=o_p(1),
$$

and

$$
\sup_{\tau\in\Upsilon}|R^w_{sfe,2,0}(u,\tau)|=o_p(1).
$$

Therefore,

$$
\begin{aligned}
L^w_{2,n}(u,\tau)=&\frac{1}{2}u'Q_{sfe}(\tau)u-\sum_{s\in\mathcal{S}}q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1+f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D^w_n(s)}{\sqrt{n}}\\
&+R^w_{sfe,2,1}(u,\tau)+R^w_{sfe,2,0}(u,\tau)+h^w_{2,1}(\tau)+h^w_{2,0}(\tau).
\end{aligned}
$$

Combining (1.11.15), (1.11.17), (1.11.18), and (1.11.19), we have

$$L^w_{sfe,n}(u,\tau) = -u'\tilde{W}^w_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}u + \tilde{R}^w_{sfe,n}(u,\tau) + h^w_{sfe,n}(\tau),$$

where

$$\begin{aligned}
&W^w_{sfe,n}(\tau)\\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i\Big[\iota_1 A_i 1\{S_i=s\}\eta_{i,1}(s,\tau) + \iota_0(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)\Big]\\
&+\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i\Big\{\iota_2(m_1(s,\tau)-m_0(s,\tau)) + q(\tau)\Big[f_1(q_1(\tau)|s)\pi\iota_1 + f_0(q_0(\tau)|s)(1-\pi)\iota_0\Big]\Big\}\\
&\times(A_i-\pi)1\{S_i=s\} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\iota_1\pi m_1(S_i,\tau) + \iota_0(1-\pi)m_0(S_i,\tau)),
\end{aligned}$$

$$h^w_{sfe,n}(\tau) = h^w_{1,1}(\tau) + h^w_{1,0}(\tau) + h^w_{2,1}(\tau) + h^w_{2,0}(\tau),$$

and

$$\sup_{\tau\in\Upsilon}|\tilde{R}^w_{sfe,n}(u,\tau)| = o_p(1).$$

In addition, by Lemma 1.11.4, $\sup_{\tau\in\Upsilon}|W^w_{sfe,n}(\tau)| = O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^w_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1}W^w_{sfe,n}(\tau) + R^w_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}||R^w_{sfe,n}(\tau)|| = o_p(1).$$

This concludes Step 1.

Step 2. We now focus on the second element of $\hat{\beta}^w_{sfe}(\tau)$. From Step 1, we know that

$$\sqrt{n}(\hat{\beta}^w_{sfe,1}(\tau) - q(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i\mathcal{J}_i(s,\tau) + R^w_{sfe,n,1}(\tau),$$

where

$$\begin{aligned}\mathcal{J}_i(s,\tau) = &\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] \\ &+ \left\{\left(\frac{1-\pi}{\pi f_1(q_1(\tau))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau)-m_0(s,\tau)) \right. \\ &\left. + q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))} - \frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right]\right\}(A_i-\pi)1\{S_i=s\} \\ &+ \left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i=s\}\end{aligned}$$

and

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,n,1}(\tau)| = o_p(1).$$

By (1.11.14), we have

$$\sqrt{n}(\hat\beta_{sfe,1}(\tau) - q(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n \mathcal{J}_i(s,\tau) + R_{sfe,n,1}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}|R_{sfe,n,1}(\tau)| = o_p(1).$$

Taking the difference of the above two equations, we have

$$\sqrt{n}(\hat\beta^w_{sfe,1}(\tau) - \hat\beta_{sfe,1}(\tau)) = \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n (\xi_i-1)\mathcal{J}_i(s,\tau) + R^w(\tau),$$

where

$$\sup_{\tau\in\Upsilon}|R^w(\tau)| = o_p(1).$$

Lemma 1.11.5 shows that, conditionally on data,

$$\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n (\xi_i-1)\mathcal{J}_i(s,\tau) \rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau),$$

where $\tilde{\mathcal{B}}_{sfe}(\tau)$ is a Gaussian process with covariance kernel

$$
\begin{aligned}
&\tilde{\Sigma}_{sfe}(\tau_1,\tau_2)\\
=&\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}+\frac{\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\\
&+\mathbb{E}\pi(1-\pi)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right)\\
&+q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))}-\frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right)\Bigg]\\
&\times\left[(m_1(S,\tau_2)-m_0(S,\tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right)+q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))}-\frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\right]\\
&+\mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}-\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}-\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right]. \qquad (1.11.20)
\end{aligned}
$$

This concludes the proof for the SFE estimator.

1.11.4 Proof of Theorem 1.11.3

Recall the definition of $\tilde{\beta}(\tau)=(\tilde{\beta}_0(\tau),\tilde{\beta}_1(\tau))'$ in the proof of Theorem 1.11.1. Let $u_0=\sqrt{n}(b_0-\tilde{\beta}_0(\tau))$, $u_1=\sqrt{n}(b_1-\tilde{\beta}_1(\tau))$ and $u=(u_0,u_1)'\in\Re^2$. Then,

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau)-\tilde{\beta}(\tau))=\arg\min_u L^*_{sfe,n}(u,\tau),$$

where

$$L^*_{sfe,n}(u,\tau)=\sum_{i=1}^n\left[\rho_\tau(Y_i^*-\dot{\breve{A}}_i^{*\prime}(\tilde{\beta}(\tau)+\frac{u}{\sqrt{n}}))-\rho_\tau(Y_i^*-\breve{A}_i^{*\prime}\tilde{\beta}(\tau))\right]$$

and $\breve{A}_i^*=(1,A_i^*-\pi)'$. Following the proof of Theorem 1.11.1, we divide the current proof into two steps. In the first step, we show that there exist

$$g^*_{sfe,n}(u,\tau)=-u'W^*_{sfe,n}(\tau)+\frac{1}{2}u'Q_{sfe}(\tau)u$$

and $h^*_{sfe,n}(\tau)$ independent of u such that for each u

$$\sup_{\tau\in\Upsilon}|L^*_{sfe,n}(u,\tau)-g^*_{sfe,n}(u,\tau)-h^*_{sfe,n}(\tau)|\xrightarrow{p}0.$$

In addition, we show that $\sup_{\tau\in\Upsilon}||W^*_{sfe,n}(\tau)|| = O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau) - \tilde{\beta}(\tau)) = [Q_{sfe}(\tau)]^{-1}W^*_{sfe,n}(\tau) + R^*_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}||R^*_{sfe,n}(\tau)|| = o_p(1).$$

In the second step, we show that, conditionally on data,

$$\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau)) \rightsquigarrow \mathcal{B}_{sfe}(\tau).$$

Step 1. Following Step 1 in the proof of Theorem 1.11.1, we have

$$L^*_{sfe,n}(u,\tau) \equiv -L^*_{1,n}(u,\tau) + L^*_{2,n}(u,\tau),$$

where

$$\begin{aligned}
&L^*_{1,n}(u,\tau)\\
=&\sum_{i=1}^{n}\sum_{s\in\mathcal{S}} A^*_i 1\{S^*_i = s\}\left(\frac{u_0}{\sqrt{n}} + (1-\hat{\pi}^*(s))\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^*(s))q(\tau)\right)(\tau - 1\{Y^*_i \le q_1(\tau)\})\\
&+\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}(1-A^*_i)1\{S^*_i = s\}\left(\frac{u_0}{\sqrt{n}} - \hat{\pi}^*(s)\frac{u_1}{\sqrt{n}} + (\pi - \hat{\pi}^*(s))q(\tau)\right)(\tau - 1\{Y^*_i \le q_0(\tau)\})\\
\equiv&L^*_{1,1,n}(u,\tau) + L^*_{1,0,n}(u,\tau),
\end{aligned}$$

$$\begin{aligned}
&L^*_{2,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=1}^{n} A^*_i 1\{S^*_i = s\}\int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E^*_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} (1\{Y^*_i \le q_1(\tau)+v\} - 1\{Y^*_i \le q_1(\tau)\})\,dv\\
&+\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(1-A^*_i)1\{S^*_i = s\}\int_0^{\frac{u'\iota_0}{\sqrt{n}} - \frac{E^*_n(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} (1\{Y^*_i \le q_0(\tau)+v\} - 1\{Y^*_i \le q_0(\tau)\})\,dv\\
\equiv&L^*_{2,1,n}(u,\tau) + L^*_{2,0,n}(u,\tau),
\end{aligned}$$

and $E^*_n(s) = \sqrt{n}(\hat{\pi}^*(s) - \pi)$.

Step 1.1. Recall that $\iota_1 = (1, 1-\pi)'$ and $\iota_0 = (1, -\pi)'$. In addition, $\hat{\pi}^*(s) - \pi = \frac{D_n^*(s)}{n^*(s)}$. Then,

$$\begin{aligned}
&L^*_{1,1,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}} \frac{u'\iota_1}{\sqrt{n}} \sum_{i=1}^n \left[A_i^* 1\{S_i^* = s\}\eta^*_{i,1}(s,\tau) + (A_i^* - \pi)1\{S_i^* = s\}m_1(s,\tau) + \pi 1\{S_i^* = s\}m_1(s,\tau)\right]\\
&- \sum_{s\in\mathcal{S}} \frac{u_1 D_n^*(s)\pi m_1(s,\tau)}{\sqrt{n}} + h^*_{1,1}(\tau) + R^*_{sfe,1,1}(u,\tau), \qquad (1.11.21)
\end{aligned}$$

where $\eta^*_{i,1}(s,\tau) = (\tau - 1\{Y_i^*(1) \le q_1(\tau)\}) - m_1(s,\tau)$,

$$h^*_{1,1}(\tau) = \sum_{s\in\mathcal{S}} (\pi - \hat{\pi}^*(s)) q(\tau) \left(\sum_{i=1}^n A_i^* 1\{S_i^* = s\}(\tau - 1\{Y_i^* \le q_1(\tau)\}) \right),$$

and

$$R^*_{sfe,1,1}(u,\tau) = -\sum_{s\in\mathcal{S}} \frac{u_1 D_n^*(s)}{\sqrt{n}n^*(s)} \left\{ \sum_{i=1}^n A_i^* 1\{S_i^* = s\}\eta^*_{i,1}(s,\tau) + (A_i^* - \pi)1\{S_i^* = s\}m_1(s,\tau) \right\}. \qquad (1.11.22)$$

Note that

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n (A_i^* - \pi)1\{S_i^* = s\}m_1(s,\tau)| = \sup_{s\in\mathcal{S},\tau\in\Upsilon} |D_n^*(s)m_1(s,\tau)| = O_p(\sqrt{n}).$$

In addition, Lemma 1.10.5 shows

$$\sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n A_i^* 1\{S_i^* = s\}\eta^*_{i,1}(s,\tau)| = O_p(\sqrt{n(s)}).$$

Therefore, we have

$$\begin{aligned}
&\sup_{\tau\in\Upsilon} |R^*_{sfe,1,1}(u,\tau)|\\
\le&\sum_{s\in\mathcal{S}} \sup_{s\in\mathcal{S}} \left|\frac{u_1 D_n^*(s)}{\sqrt{n}n^*(s)}\right| \left[\sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n A_i^* 1\{S_i^* = s\}\eta^*_{i,1}(s,\tau)| + \sup_{s\in\mathcal{S},\tau\in\Upsilon} |\sum_{i=1}^n (A_i^* - \pi)1\{S_i^* = s\}m_1(s,\tau)|\right]\\
=&O_p(1/\sqrt{n}).
\end{aligned}$$

Similarly, we have

$$
\begin{aligned}
&L^*_{1,0,n}(u,\tau) \\
=&\sum_{s\in\mathcal{S}}\frac{u'\iota_0}{\sqrt{n}}\sum_{i=1}^{n}\left[(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)-(A_i^*-\pi)1\{S_i^*=s\}m_0(s,\tau)+(1-\pi)1\{S_i^*=s\}m_0(s,\tau)\right] \\
&-\sum_{s\in\mathcal{S}}\frac{u_1D_n^*(s)(1-\pi)m_0(s,\tau)}{\sqrt{n}}+h^*_{1,0}(\tau)+R^*_{sfe,1,0}(u,\tau), \qquad (1.11.23)
\end{aligned}
$$

where

$$
h^*_{1,0}(\tau)=\sum_{s\in\mathcal{S}}(\pi-\hat{\pi}^*(s))q(\tau)\left(\sum_{i=1}^{n}(1-A_i^*)1\{S_i^*=s\}(\tau-1\{Y_i^*\le q_0(\tau)\})\right),
$$

and

$$
R^*_{sfe,1,0}(u,\tau)=-\sum_{s\in\mathcal{S}}\frac{u_1D_n^*(s)}{\sqrt{n}n^*(s)}\left\{\sum_{i=1}^{n}(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)-(A_i^*-\pi)1\{S_i^*=s\}m_0(s,\tau)\right\} \qquad (1.11.24)
$$

such that

$$
\sup_{\tau\in\Upsilon}|R^*_{sfe,1,0}(u,\tau)|=O_p(1/\sqrt{n}).
$$

Therefore,

$$
\begin{aligned}
L^*_{1,n}(u,\tau)=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[u'\iota_1A_i^*1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)+u'\iota_0(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)\right] \\
&+\sum_{s\in\mathcal{S}}u'\iota_2\frac{D_n^*(s)}{\sqrt{n}}\left(m_1(s,\tau)-m_0(s,\tau)\right) \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(u'\iota_1\pi m_1(S_i^*,\tau)+u'\iota_0(1-\pi)m_0(S_i^*,\tau)) \\
&+R^*_{sfe,1,1}(u,\tau)+R^*_{sfe,1,0}(u,\tau)+h_{1,1}(\tau)+h_{1,0}(\tau).
\end{aligned}
$$

Furthermore, by Lemma 1.11.6, we have

$$L^*_{2,1,n}(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D^*_n(s)u'\iota_1}{\sqrt{n}}q(\tau) + h^*_{2,1}(\tau) + R^*_{sfe,2,1}(u,\tau) \tag{1.11.25}$$

and

$$L^*_{2,0,n}(u,\tau) = \frac{(1-\pi)f_0(q_0(\tau))}{2}(u'\iota_0)^2 - \sum_{s\in\mathcal{S}} f_0(q_0(\tau)|s)\frac{(1-\pi)D^*_n(s)u'\iota_0}{\sqrt{n}}q(\tau) + h^*_{2,0}(\tau) + R^*_{sfe,2,0}(u,\tau), \tag{1.11.26}$$

where

$$h^*_{2,1}(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2}p(s)(E^*_n(s))^2 q^2(\tau),$$

$$h^*_{2,0}(\tau) = \sum_{s\in\mathcal{S}} \frac{(1-\pi)f_0(q_0(\tau)|s)}{2}p(s)(E^*_n(s))^2 q^2(\tau),$$

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,2,1}(u,\tau)| = o_p(1),$$

and

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,2,0}(u,\tau)| = o_p(1).$$

Therefore,

$$\begin{aligned} L^*_{2,n}(u,\tau) =& \frac{1}{2}u'Q_{sfe}(\tau)u - \sum_{s\in\mathcal{S}} q(\tau)\left[f_1(q_1(\tau)|s)\pi u'\iota_1 + f_0(q_0(\tau)|s)(1-\pi)u'\iota_0\right]\frac{D^*_n(s)}{\sqrt{n}} \\ &+ R^*_{sfe,2,1}(u,\tau) + R^*_{sfe,2,0}(u,\tau) + h^*_{2,1}(\tau) + h^*_{2,0}(\tau). \end{aligned}$$

Combining (1.11.21), (1.11.23), (1.11.25), and (1.11.26), we have

$$L^*_{sfe,n}(u,\tau) = -u'W^*_{sfe,n}(\tau) + \frac{1}{2}u'Q_{sfe}u + \tilde{R}^*_{sfe,n}(u,\tau) + h^*_{sfe,n}(\tau),$$

where

$$
\begin{aligned}
&W^*_{sfe,n}(\tau) \\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\Big[\iota_1 A_i^* 1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)+\iota_0(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)\Big] \\
&+\sum_{s\in\mathcal{S}}\Big\{\iota_2(m_1(s,\tau)-m_0(s,\tau))+q(\tau)\Big[f_1(q_1(\tau)|s)\pi\iota_1+f_0(q_0(\tau)|s)(1-\pi)\iota_0\Big]\Big\}\frac{D_n^*(s)}{\sqrt{n}} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\iota_1\pi m_1(S_i^*,\tau)+\iota_0(1-\pi)m_0(S_i^*,\tau)),
\end{aligned}
$$

$$h^*_{sfe,n}(\tau)=h^*_{1,1}(\tau)+h^*_{1,0}(\tau)+h^*_{2,1}(\tau)+h^*_{2,0}(\tau),$$

and

$$\sup_{\tau\in\Upsilon}|\tilde{R}^*_{sfe,n}(u,\tau)|=o_p(1).$$

By Lemma 1.11.7, $\sup_{\tau\in\Upsilon}|W^*_{sfe,n}(\tau)|=O_p(1)$. Then, by Kato (2009, Theorem 2), we have

$$\sqrt{n}(\hat{\beta}^*_{sfe}(\tau)-\tilde{\beta}(\tau))=[Q_{sfe}(\tau)]^{-1}W^*_{sfe,n}(\tau)+R^*_{sfe,n}(\tau),$$

where

$$\sup_{\tau\in\Upsilon}||R^*_{sfe,n}(\tau)||=o_p(1).$$

This concludes Step 1.

Step 2. We now focus on the second element of $\hat{\beta}^*_{sfe}(\tau)$. From Step 1, we know that

$$
\begin{aligned}
&\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau)-q(\tau)) \\
=&\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i^*1\{S_i^*=s\}\eta^*_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i^*)1\{S_i^*=s\}\eta^*_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] \\
&+\sum_{s\in\mathcal{S}}\left\{\left(\frac{1-\pi}{\pi f_1(q_1(\tau))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau)-m_0(s,\tau))+q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))}-\frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right]\right\} D \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{m_1(S_i^*,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i^*,\tau)}{f_0(q_0(\tau))}\right)+R^*_{sfe,n,1}(\tau) \\
\equiv&W^*_{sfe,n,1}(\tau)+W^*_{sfe,n,2}(\tau)+W^*_{sfe,n,3}(\tau)+R^*_{sfe,n,1}(\tau),
\end{aligned}
$$

where

$$\sup_{\tau \in \Upsilon} |R^*_{sfe,n,1}(\tau)| = o_p(1).$$

By (1.10.8), we have

$$\begin{aligned}
&\sqrt{n}(\hat{q}(\tau) - q(\tau)) \\
=&\frac{1}{\sqrt{n}} \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \left[\frac{A_i 1\{S_i = s\} \eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{(1-A_i) 1\{S_i = s\} \eta_{i,0}(s,\tau)}{(1-\pi) f_0(q_0(\tau))} \right] \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))} \right) + R_{ipw,n}(\tau) \\
\equiv&\mathcal{W}_{n,1}(\tau) + \mathcal{W}_{n,2}(\tau) + R_{ipw,n}(\tau),
\end{aligned}$$

where

$$\sup_{\tau \in \Upsilon} |R_{ipw,n}(\tau)| = o_p(1).$$

Taking the difference of the above two equations, we have

$$\sqrt{n}(\hat{\beta}^*_{sfe,1}(\tau) - \hat{q}(\tau)) = (W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau)) + W^*_{sfe,n,2}(\tau) + (W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)) + R^*(\tau), \tag{1.11.27}$$

where

$$\sup_{\tau \in \Upsilon} |R^*(\tau)| = o_p(1).$$

Lemma 1.11.7 shows that, conditionally on data,

$$(W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau)), W^*_{sfe,n,2}(\tau), (W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ are three independent Gaussian processes and $\sum_{j=1}^{3} \mathcal{B}_j(\tau) \stackrel{d}{=} \mathcal{B}_{sfe}(\tau)$. This concludes the proof.

1.11.5 Technical Lemmas

Lemma 1.11.1. *Let $W_{sfe,n,j}(\tau)$, $j=1,2,3$ be defined as in (1.11.14). If Assumptions in Theorem 1.11.1 hold, then uniformly over $\tau \in \Upsilon$,*

$$(W_{sfe,n,1}(\tau), W_{sfe,n,2}(\tau), W_{sfe,n,3}(\tau)) \rightsquigarrow (\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau)),$$

where $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are three independent two-dimensional Gaussian process with covariance kernels $\Sigma_{sfe,1}(\tau_1,\tau_2)$, $\Sigma_{sfe,2}(\tau_1,\tau_2)$, and $\Sigma_{sfe,3}(\tau_1,\tau_2)$, respectively. The expressions for the three kernels are derived in the proof below.

Proof. The proofs of weak convergence and the independence among $(\mathcal{B}_{sfe,1}(\tau), \mathcal{B}_{sfe,2}(\tau), \mathcal{B}_{sfe,3}(\tau))$ are similar to that in Lemma 1.10.2, and thus, are omitted. In the following, we focus on deriving the covariance kernels.

First, similar to the argument in the proof of Lemma 1.10.2,

$$W_{sfe,n,1}(\tau) \stackrel{d}{=} \iota_1 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+1}^{N(s)+n_1(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,1}(s,\tau) + \iota_0 \sum_{s\in\mathcal{S}} \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)} \frac{1}{\sqrt{n}} \tilde{\eta}_{i,0}(s,\tau).$$

Therefore,

$$\begin{aligned}\Sigma_1(\tau_1,\tau_2) =& \pi[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)]\iota_1\iota_1' \\ &+ (1-\pi)[\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)]\iota_0\iota_0'.\end{aligned}$$

For $W_{sfe,n,2}(\tau)$, we have

$$\begin{aligned}\Sigma_2(\tau_1,\tau_2) =& \mathbb{E}\gamma(S)\left[\iota_2(m_1(S,\tau_1) - m_0(S,\tau_1)) + q(\tau_1)\Big(f_1(q_1(\tau_1)|S)\pi\iota_1 + f_0(q_0(\tau_1)|S)(1-\pi)\iota_0\Big)\right] \\ &\times \left[\iota_2(m_1(S,\tau_2) - m_0(S,\tau_2)) + q(\tau_2)\Big(f_1(q_1(\tau_2)|S)\pi\iota_1 + f_0(q_0(\tau_2)|S)(1-\pi)\iota_0\Big)\right]'.\end{aligned}$$

Next, we have

$$\Sigma_3(\tau_1,\tau_2) = \mathbb{E}(\iota_1\pi m_1(S,\tau_1) + \iota_0(1-\pi)m_0(S,\tau_1))(\iota_1\pi m_1(S,\tau_2) + \iota_0(1-\pi)m_0(S,\tau_2))'.$$

In addition,

$$[Q_{sfe}(\tau)]^{-1} = \begin{pmatrix} \frac{1-\pi}{f_0(q_0(\tau))} + \frac{\pi}{f_1(q_1(\tau))} & \frac{1}{f_1(q_1(\tau))} - \frac{1}{f_0(q_0(\tau))} \\ \frac{1}{f_1(q_1(\tau))} - \frac{1}{f_0(q_0(\tau))} & \frac{1}{(1-\pi)f_0(q_0(\tau))} + \frac{1}{\pi f_1(q_1(\tau))} \end{pmatrix}.$$

Therefore,

$$
\begin{aligned}
&\Sigma(\tau_1,\tau_2)\\
=&\Bigg\{\frac{1}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))}\left[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)\right]\begin{pmatrix}\pi^2 & \pi\\ \pi & 1\end{pmatrix}\\
&+\frac{1}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))}\left[\min(\tau_1,\tau_2)-\tau_1\tau_2-\mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)\right]\begin{pmatrix}(1-\pi)^2 & \pi-1\\ \pi-1 & 1\end{pmatrix}\Bigg\}\\
&+\Bigg\{\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1)-m_0(S,\tau_1))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_1))}+\frac{1-\pi}{f_1(q_1(\tau_1))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_1))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\end{pmatrix}+q(\tau_1)\frac{f_1(q_1(\tau_1)|S)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\1\end{pmatrix}\\
&+q(\tau_1)\frac{f_0(q_0(\tau_1)|S)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\times\Bigg[(m_1(S,\tau_2)-m_0(S,\tau_2))\begin{pmatrix}\frac{\pi}{f_0(q_0(\tau_2))}+\frac{1-\pi}{f_1(q_1(\tau_2))}\\ \frac{1-\pi}{\pi f_1(q_1(\tau_2))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\end{pmatrix}\\
&+q(\tau_2)\frac{f_1(q_1(\tau_2)|S)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\1\end{pmatrix}+q(\tau_2)\frac{f_0(q_0(\tau_2)|S)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\Bigg\}\\
&+\Bigg\{\mathbb{E}\Bigg[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))}\begin{pmatrix}\pi\\1\end{pmatrix}+\frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]\Bigg[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))}\begin{pmatrix}\pi\\1\end{pmatrix}+\frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\begin{pmatrix}1-\pi\\-1\end{pmatrix}\Bigg]'\Bigg\}.
\end{aligned}
$$

□

Lemma 1.11.2. *Recall the definition of $R^w_{sfe,1,1}(u,\tau)$ in (1.11.16). If Assumptions 1 and 2 hold, then*

$$\sup_{\tau\in\Upsilon}|R^w_{sfe,1,1}(u,\tau)|=o_p(1).$$

Proof. We divide the proof into two steps. In the first step, we show that $\sup_{s\in\mathcal{S}}|D^w_n(s)|=O_p(\sqrt{n})$. In the second step, we show that

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|\sum_{i=1}^n\xi_iA_i1\{S_i=s\}\eta_{i,1}(s,\tau)|=O_p(\sqrt{n}). \tag{1.11.28}$$

Then,

$$
\begin{aligned}
&\sup_{\tau\in\Upsilon}|R^w_{sfe,1,1}(u,\tau)|\\
\leq&\sum_{s\in\mathcal{S}}\frac{|u_1|}{n^w(s)}\sup_{s\in\mathcal{S}}\left|\frac{D^w_n(s)}{\sqrt{n}}\right|\left[\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{i=1}^n\xi_iA_i1\{S_i=s\}\eta_{i,1}(s,\tau)\right|+\sup_{s\in\mathcal{S}}|D^w_n(s)|\right]\\
=&O_p(1/\sqrt{n}),
\end{aligned}
$$

as $n^w(s)/n \xrightarrow{p} p(s) > 0$.

Step 1. Because

$$\sup_{s\in\mathcal{S}} |D_n(s)| = O_p(\sqrt{n}),$$

we only need to bound the difference $D_n^w(s) - D_n(s)$. Note that

$$n(s)^{-1/2} D_n^w(s) - n(s)^{-1/2} D_n(s) = n^{-1/2} \sum_{i=1}^{n} (\xi_i - 1)(A_i - \pi) 1\{S_i = s\}. \tag{1.11.29}$$

We aim to prove that, if $n(s) \to \infty$ and $D_n(s)/n(s) = o_p(1)$, then conditionally on data, for $s \in \mathcal{S}$,

$$n(s)^{-1/2} \sum_{i=1}^{n} (\xi_i - 1)(A_i - \pi) 1\{S_i = s\} \rightsquigarrow N(0, \pi(1-\pi)) \tag{1.11.30}$$

and they are independent across $s \in \mathcal{S}$. The independence is straightforward because

$$\frac{1}{n(s)} \sum_{i=1}^{n} (\xi_i - 1)^2 (A_i - \pi)^2 1\{S_i = s\} 1\{S_i = s'\} = 0 \quad \text{for} \quad s \neq s'.$$

For the limiting distribution, let $\mathcal{D}_n = \{Y_i, A_i, S_i\}_{i=1}^n$ denote data. According to the Lindeberg-Feller central limit theorem, (1.11.30) holds because (1)

$$\begin{aligned}
n(s)^{-1} \sum_{i=1}^{n} \mathbb{E}[(\xi_i - 1)^2 (A_i - \pi)^2 1\{S_i = s\} | \mathcal{D}_n] =& n(s)^{-1} \sum_{i=1}^{n} (A_i - 2A_i\pi + \pi^2) 1\{S_i = s\} \\
=& n(s)^{-1} \sum_{i=1}^{n} (A_i - \pi - 2(A_i - \pi)\pi + \pi - \pi^2) 1\{S_i = s\} \\
=& \frac{1 - 2\pi}{n(s)} D_n(s) + \pi(1-\pi) \\
\xrightarrow{p}& \pi(1-\pi),
\end{aligned}$$

and (2) for every $\varepsilon > 0$,

$$\begin{aligned}
& n(s)^{-1} \sum_{i=1}^{n} (A_i - \pi)^2 1\{S_i = s\} \mathbb{E}\left[(\xi_i - 1)^2 1\{|\xi_i - 1| (A_i - \pi)^2 1\{S_i = s\} > \varepsilon \sqrt{n(s)}\} | \mathcal{D}_n \right] \\
\leq & 4\mathbb{E}(\xi_i - 1)^2 1\{2|\xi_i - 1| \geq \varepsilon \sqrt{n(s)}\} \to 0,
\end{aligned}$$

where we use the fact that $|A_i - \pi|1\{S_i = s\} \leq 2$ and $n(s) \to \infty$. This concludes the proof of Step 1.

Step 2. By the same rearrangement argument and the fact that $\{\xi_i\}_{i=1}^n \perp\!\!\!\perp \mathcal{D}_n$, we have

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)\right| \stackrel{d}{=} \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i \tilde{\eta}_{i,1}(s,\tau)\right|.$$

Let $\Gamma_{n,1}(s,t,\tau) = \sum_{i=1}^{\lfloor nt \rfloor} \frac{\xi_i \tilde{\eta}_{i,1}(s,\tau)}{\sqrt{n}}$ and $\mathcal{F} = \{\xi_i \tilde{\eta}_{i,1}(s,\tau) : \tau \in \Upsilon, s \in \mathcal{S}\}$ with envelope $F_i = C\xi_i$ and $||F_i||_{P,2} < \infty$. By Lemma 1.10.1 and van der Vaart and Wellner (1996, Theorem 2.14.1), for any $\varepsilon > 0$, we can choose M sufficiently large such that

$$\begin{aligned}\mathbb{P}(\sup_{0<t\leq 1, \tau\in\Upsilon, s\in\mathcal{S}} |\Gamma_{n,1}(s,t,\tau)| \geq M) \leq & \frac{270\mathbb{E}\sup_{\tau\in\Upsilon, s\in\mathcal{S}} |\Gamma_{n,1}(s,1,\tau)|}{M} \\ = & \frac{270\mathbb{E}\sqrt{n}||\mathbb{P}_n - \mathbb{P}||_{\mathcal{F}}}{M} \lesssim \frac{J(1,\mathcal{F})||F_i||_{P,2}}{M} < \varepsilon.\end{aligned}$$

Therefore,

$$\sup_{0<t\leq 1, \tau\in\Upsilon, s\in\mathcal{S}} |\Gamma_{n,1}(s,t,\tau)| = O_p(1)$$

and

$$\begin{aligned}\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left|\frac{1}{n}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau)\right| \stackrel{d}{=} & \sup_{\tau\in\Upsilon, s\in\mathcal{S}} \frac{1}{\sqrt{n}}\left|\Gamma_{n,1}\left(s, \frac{N(s)+n_1(s)}{n}, \tau\right) - \Gamma_{n,1}\left(s, \frac{N(s)}{n}, \tau\right)\right| \\ = & O_p(1/\sqrt{n}). \qquad (1.11.31)\end{aligned}$$

This concludes the proof of Step 2. □

Lemma 1.11.3. *If Assumptions 1 and 2 hold, then 1.11.18 and 1.11.19 hold.*

Proof. We focus on (1.11.18). Note that

$$\begin{aligned}& L^w_{2,1,n}(u,\tau) \\ = & \sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\} \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E^w_n(s)}{\sqrt{n}}\left(q(\tau) + \frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i(1) \leq q_1(\tau) + v\} - 1\{Y_i(1) \leq q_1(\tau)\}\right) dv \\ = & \sum_{s\in\mathcal{S}}\sum_{i=1}^n \xi_i A_i 1\{S_i = s\}[\phi_i(u,\tau,s,E^w_n(s)) - \mathbb{E}\phi_i(u,\tau,s,E^w_n(s)|S_i = s)]\end{aligned}$$

$$+\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i A_i 1\{S_i=s\}\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i=s), \tag{1.11.32}$$

where by Lemma 1.11.2, $E_n^w(s)=\sqrt{n}(\hat{\pi}^w(s)-\pi)=\frac{n}{n^w(s)}\frac{D_n^w(s)}{\sqrt{n}}=O_p(1)$,

$$\phi_i(u,\tau,s,e)=\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i(1)\le q_1(\tau)+v\}-1\{Y_i(1)\le q_1(\tau)\}\right)dv,$$

and $\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i=s)$ is interpreted as $\mathbb{E}(\phi_i(u,\tau,s,e)|S_i=s)$ with e being evaluated at $E_n^w(s)$.

For the first term on the RHS of (1.11.32), by the rearrangement argument in Lemma 1.10.2, we have

$$\begin{aligned}&\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i A_i 1\{S_i=s\}[\phi_i(u,\tau,s,E_n^w(s))-\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i=s)]\\ \stackrel{d}{=}&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\xi_i[\phi_i^s(u,\tau,s,E_n^w(s))-\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))],\end{aligned}$$

where

$$\phi_i^s(u,\tau,s,e)=\int_0^{\frac{u'\iota_1}{\sqrt{n}}-\frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)}\left(1\{Y_i^s(1)\le q_1(\tau)+v\}-1\{Y_i^s(1)\le q_1(\tau)\}\right)dv.$$

Similar to (1.11.9), we can show that, as $n\to\infty$,

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{i=N(s)+1}^{N(s)+n_1(s)}\xi_i\left[\phi_i^s(u,\tau,s,E_n^w(s))-\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))\right]\right|=o_p(1). \tag{1.11.33}$$

For the second term in (1.11.32), we have

$$\begin{aligned}&\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\xi_i A_i 1\{S_i=s\}\mathbb{E}\phi_i(u,\tau,s,E_n^w(s)|S_i=s)\\ =&\sum_{s\in\mathcal{S}}\frac{\sum_{i=1}^n\xi_i\pi 1\{S_i=s\}}{n}n\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))+\sum_{s\in\mathcal{S}}\frac{D_n^w(s)}{n}n\mathbb{E}\phi_i^s(u,\tau,s,E_n^w(s))\\ =&\sum_{s\in\mathcal{S}}\pi p(s)\left[\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1-E_n^w(s)q(\tau))^2+o_p(1)\right]+\sum_{s\in\mathcal{S}}\frac{D_n^w(s)}{n}\left[\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1-E_n^w(s)q(\tau))^2+o_p\right.\end{aligned}$$

$$=\frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n^w(s)u'\iota_1}{\sqrt{n}}q(\tau) + h_{2,1}^w(\tau) + o_p(1), \tag{1.11.34}$$

where the $o_p(1)$ term holds uniformly over $(\tau, s) \in \Upsilon \times \mathcal{S}$. The second equality holds by the same calculation in (1.11.10) and the fact that $\sum_{i=1}^n \xi_i 1\{S_i = s\}/n \xrightarrow{p} p(s)$. The last inequality holds because $\frac{D_n^w(s)}{n} = o_p(1)$, $E_n^w(s) = \frac{n}{n^w(s)}\frac{D_n^w(s)}{\sqrt{n}} = O_p(1)$, $\frac{n}{n^w(s)} \xrightarrow{p} 1/p(s)$, and

$$h_{2,1}^w(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)(E_n^w(s))^2 q^2(\tau).$$

Combining (1.11.32)–(1.11.34), we have

$$L_{2,1,n}^w(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n^w(s)u'\iota_1}{\sqrt{n}}q(\tau) + h_{2,1}^w(\tau) + R_{sfe,2,1}^w(u,\tau),$$

where

$$h_{2,1}^w(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)(E_n^w(s))^2 q^2(\tau)$$

and

$$\sup_{\tau\in\Upsilon} |R_{sfe,2,1}^w(u,\tau)| = o_p(1).$$

This concludes the proof.

□

Lemma 1.11.4. *If Assumptions 1 and 2 hold, then* $\sup_{\tau\in\Upsilon} ||W_{sfe,n}^w(\tau)|| = O_p(1)$.

Proof. It suffices to show that

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i A_i 1\{S_i = s\}\eta_{i,1}(s,\tau) \right| = O_p(1) \tag{1.11.35}$$

$$\sup_{\tau\in\Upsilon, s\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (1 - A_i) 1\{S_i = s\}\eta_{i,0}(s,\tau) \right| = O_p(1), \tag{1.11.36}$$

$$\sup_{s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(A_i-\pi)1\{S_i=s\}\right| = O_p(1), \tag{1.11.37}$$

and

$$\sup_{\tau\in\Upsilon}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(\iota_1\pi m_1(S_i,\tau)+\iota_0(1-\pi)m_0(S_i,\tau))\right\| = O_p(1). \tag{1.11.38}$$

Note that (1.11.35) holds by the argument in step 2 in the proof of Lemma 1.11.2, (1.11.36) holds similarly, (1.11.37) holds by (1.11.29) and (1.11.30), and (1.11.38) holds by the usual maximal inequality, e.g., van der Vaart and Wellner (1996, Theorem 2.14.1). This concludes the proof. □

Lemma 1.11.5. *If Assumptions 1 and 2 hold, then conditionally on data,*

$$\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^n(\xi_i-1)\mathcal{J}_i(s,\tau)\rightsquigarrow \tilde{\mathcal{B}}_{sfe}(\tau),$$

where $\tilde{\mathcal{B}}_{sfe}(\tau)$ *is a Gaussian process with covariance kernel* $\tilde{\Sigma}_{sfe}(\cdot,\cdot)$ *defined in* (1.11.20)*.*

Proof. In order to show the weak convergence, we only need to show (1) conditional stochastic equicontinuity and (2) conditional convergence in finite dimension. We divide the proof into two steps accordingly.

Step 1. In order to show the conditional stochastic equicontinuity, it suffices to show that, for any $\varepsilon>0$, as $n\to\infty$ followed by $\delta\to 0$,

$$\mathbb{P}_\xi\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\stackrel{p}{\longrightarrow}0,$$

where $\mathbb{P}_\xi(\cdot)$ means that the probability operator is with respect to $\xi_1,\cdots,\xi_n$ and conditional on data. Note

$$\begin{aligned}
&\mathbb{E}\mathbb{P}_\xi\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_i(s,\tau_1)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
=&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_i(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon\right)\\
\leq&\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_{i,1}(s,\tau_1))\right|\geq\varepsilon/3\right)
\end{aligned}$$

$$+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2)-\mathcal{J}_{i,2}(s,\tau_1))\right|\geq\varepsilon/3\right)$$
$$+\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,3}(s,\tau_2)-\mathcal{J}_{i,3}(s,\tau_1))\right|\geq\varepsilon/3\right),$$

where

$$\mathcal{J}_{i,1}(s,\tau)=\frac{A_i1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))},$$

$$\mathcal{J}_{i,2}(s,\tau)=F_1(s,\tau)(A_i-\pi)1\{S_i=s\},$$

$$F_1(s,\tau)=\left(\frac{1-\pi}{\pi f_1(q_1(\tau))}-\frac{\pi}{(1-\pi)f_0(q_0(\tau))}\right)(m_1(s,\tau)-m_0(s,\tau))+q(\tau)\left[\frac{f_1(q_1(\tau)|s)}{f_1(q_1(\tau))}-\frac{f_0(q_0(\tau)|s)}{f_0(q_0(\tau))}\right],$$

$$\mathcal{J}_{i,3}(s,\tau)=\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i=s\}.$$

Further note that

$$\sum_{i=1}^n(\xi_i-1)\mathcal{J}_{i,1}(s,\tau)\stackrel{d}{=}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{(\xi_i-1)\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{(\xi_i-1)\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}$$

By the same argument in Claim (1) in the proof of Lemma 1.10.2, we have

$$\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_i(s,\tau_1))\right|\geq\varepsilon/3\right)$$
$$\leq\frac{3\mathbb{E}\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i-1)(\mathcal{J}_{i,1}(s,\tau_2)-\mathcal{J}_{i,1}(s,\tau_1))\right|}{\varepsilon}$$
$$\leq\frac{3\sqrt{c_2\delta\log(\frac{C}{c_1\delta})}+\frac{3C\log(\frac{C}{c_1\delta})}{\sqrt{n}}}{\varepsilon},$$

where C, $c_1<c_2$ are some positive constants that are independent of (n,ε,δ). By letting $n\to\infty$ followed by $\delta\to 0$, the RHS vanishes.

For $\mathcal{J}_{i,2}$, we note that $F_1(s,\tau)$ is Lipschitz in τ. Therefore,

$$\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,2}(s,\tau_2)-\mathcal{J}_{i,2}(s,\tau_1))\right|\geq\varepsilon/3\right)$$
$$\leq\sum_{s\in\mathcal{S}}\mathbb{P}\left(C\delta\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(A_i-\pi)1\{S_i=s\}\right|\geq\varepsilon/3\right)\to 0$$

as $n\to\infty$ followed by $\delta\to 0$, in which we use the fact that, by (1.11.30),

$$\sup_{s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(A_i-\pi)1\{S_i=s\}\right|=O_p(1).$$

Last, by the standard maximal inequality (e.g., van der Vaart and Wellner (1996, Theorem 2.14.1)) and the fact that

$$\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)$$

is Lipschitz in τ, we have, as $n\to\infty$ followed by $\delta\to 0$,

$$\mathbb{P}\left(\sup_{\tau_1,\tau_2\in\Upsilon,\tau_1<\tau_2<\tau_1+\delta,s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i-1)(\mathcal{J}_{i,3}(s,\tau_2)-\mathcal{J}_{i,3}(s,\tau_1))\right|\geq\varepsilon/3\right)\to 0$$

This concludes the proof of the conditional stochastic equicontinuity.

Step 2. We focus on the one-dimension case and aim to show that, conditionally on data, for fixed $\tau\in\Upsilon$,

$$\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}(\xi_i-1)\mathcal{J}_i(s,\tau)\rightsquigarrow\mathcal{N}(0,\tilde{\Sigma}_{sfe}(\tau,\tau)).$$

The finite-dimensional convergence can be established similarly by the Cramér-Wold device. In view of Lindeberg-Feller central limit theorem, we only need to show that (1)

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2\xrightarrow{p}\zeta_Y^2(\pi,\tau)+\tilde{\xi}_A'^2(\pi,\tau)+\xi_S^2(\pi,\tau)$$

and (2)

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2\mathbb{E}_{\xi}(\xi-1)^2 1\{|\sum_{s\in\mathcal{S}}(\xi_i-1)\mathcal{J}_i(s,\tau)|\geq\varepsilon\sqrt{n}\}\to 0.$$

(2) is obvious as $|\mathcal{J}_i(s,\tau)|$ is bounded and $\max_i|\xi_i-1|\lesssim\log(n)$ as ξ_i is sub-exponential. Next, we focus on (1). We have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}}\mathcal{J}_i(s,\tau)]^2\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\Bigg\{\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\\
&+F_1(s,\tau)(A_i-\pi)1\{S_i=s\}+\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)1\{S_i=s\}\right]\Bigg\}^2\\
\equiv&\sigma_1^2+\sigma_2^2+\sigma_3^2+2\sigma_{12}+2\sigma_{13}+2\sigma_{23},
\end{aligned}$$

where

$$\sigma_1^2=\frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]^2,$$

$$\sigma_2^2=\frac{1}{n}\sum_{s\in\mathcal{S}}F_1^2(s,\tau)\sum_{i=1}^{n}(A_i-\pi)^2 1\{S_i=s\},$$

$$\sigma_3^2=\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))}-\frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]^2,$$

$$\sigma_{12}=\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]F_1(s,\tau)(A_i-\pi)1\{S_i=s\},$$

$$\sigma_{13}=\frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}-\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))}-\frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right],$$

and

$$\sigma_{23} = \sigma_{12} = \frac{1}{n}\sum_{i=1}^{n}\sum_{s\in\mathcal{S}} F_1(s,\tau)(A_i-\pi)1\{S_i=s\}\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right].$$

For σ_1^2, we have

$$\begin{aligned}
\sigma_1^2 =& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left[\frac{A_i 1\{S_i=s\}\eta_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} - \frac{(1-A_i)1\{S_i=s\}\eta_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))}\right] \\
\stackrel{d}{=}& \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}^2(s,\tau)}{\pi^2 f_1^2(q_1(\tau))} + \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}^2(s,\tau)}{(1-\pi)^2 f_0^2(q_0(\tau))} \\
\xrightarrow{p}& \frac{\tau(1-\tau)-\mathbb{E}m_1^s(S,\tau)}{\pi f_1^2(q_1(\tau))} + \frac{\tau(1-\tau)-\mathbb{E}m_0^s(S,\tau)}{(1-\pi)f_0^2(q_0(\tau))} = \zeta_Y^2(\pi,\tau),
\end{aligned}$$

where the second equality holds due to the rearrangement argument in Lemma 1.10.2 and the convergence in probability holds due to uniform convergence of the partial sum process.

For σ_2^2, by Assumption 1,

$$\sigma_2^2 = \frac{1}{n}\sum_{s\in\mathcal{S}} F_1^2(s,\tau)(D_n(s) - 2\pi D_n(s) + \pi(1-\pi)1\{S_i=s\}) \xrightarrow{p} \pi(1-\pi)\mathbb{E}F_1^2(S_i,\tau) = \tilde{\xi}_A'^2(\pi,\tau).$$

For σ_3^2, by the law of large number,

$$\sigma_3^2 \xrightarrow{p} \mathbb{E}\left[\left(\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right)\right]^2 = \xi_S^2(\pi,\tau).$$

For σ_{12}, we have

$$\begin{aligned}
\sigma_{12} =& \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=1}^{n}\frac{A_i 1\{S_i=s\}\eta_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=1}^{n}\frac{(1-A_i)1\{S_i=s\}\eta_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \\
\stackrel{d}{=}& \frac{1}{n}\sum_{s\in\mathcal{S}}(1-\pi)F_1(s,\tau)\sum_{i=N(s)+1}^{N(s)+n_1(s)}\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \frac{1}{n}\sum_{s\in\mathcal{S}}\pi F_1(s,\tau)\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))} \xrightarrow{p} 0,
\end{aligned}$$

where the last convergence holds because by Lemma 1.10.2,

$$\frac{1}{n}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\eta}_{i,1}(s,\tau) \xrightarrow{p} 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\tilde{\eta}_{i,0}(s,\tau) \xrightarrow{p} 0.$$

By the same argument, we can show that

$$\sigma_{13} \xrightarrow{p} 0.$$

Last, for σ_{23}, by Assumption 1,

$$\sigma_{23} = \sum_{s\in\mathcal{S}} F_1(s,\tau)\left[\left(\frac{m_1(s,\tau)}{f_1(q_1(\tau))} - \frac{m_0(s,\tau)}{f_0(q_0(\tau))}\right)\right]\frac{D_n(s)}{n} \xrightarrow{p} 0.$$

Therefore, we have

$$\frac{1}{n}\sum_{i=1}^{n}[\sum_{s\in\mathcal{S}} \mathcal{J}_i(s,\tau)]^2 \xrightarrow{p} \zeta_Y^2(\pi,\tau) + \tilde{\xi}_A'^2(\pi,\tau) + \xi_S^2(\pi,\tau).$$

□

Lemma 1.11.6. *Recall $R^*_{sfe,2,1}(u,\tau)$ and $R^*_{sfe,2,0}(u,\tau)$ defined in* (1.11.25) *and* (1.11.26)*, respectively. If Assumptions in Theorem 1.5.1 hold, then* (1.11.25) *and* (1.11.26) *hold and*

$$\sup_{\tau\in\Upsilon}|R^*_{sfe,2,1}(u,\tau)| = o_p(1) \quad \textit{and} \quad \sup_{\tau\in\Upsilon}|R^*_{sfe,2,0}(u,\tau)| = o_p(1).$$

Proof. We focus on (1.11.25). Following the definition of M_{ni} in the proof of Lemma 1.10.5 and the argument in the Step 1.2 of the proof of Theorem 1.11.1, we have

$$\begin{aligned}
&L^*_{2,1,n}(u,\tau)\\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{E_n^*(s)}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i^s(1)\leq q_1(\tau)+v\} - 1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv\\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,E_n^*(s))\right] + \sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\mathbb{E}\phi_i(u,\tau,s,E_n^*(s)),
\end{aligned} \tag{1.11.39}$$

where $E_n^*(s) = \sqrt{n}(\hat{\pi}^*(s)-\pi) = \frac{n}{n^*(s)}\frac{D_n^*(s)}{\sqrt{n}} = O_p(1)$,

$$\phi_i(u,\tau,s,e) = \int_0^{\frac{u'\iota_1}{\sqrt{n}} - \frac{e}{\sqrt{n}}\left(q(\tau)+\frac{u_1}{\sqrt{n}}\right)} \left(1\{Y_i^s(1)\leq q_1(\tau)+v\} - 1\{Y_i^s(1)\leq q_1(\tau)\}\right)dv,$$

and $\mathbb{E}\phi_i(u,\tau,s,E_n^*(s))$ is interpreted as $\mathbb{E}\phi_i(u,\tau,s,e)$ with e being evaluated at $E_n^*(s)$.

For the first term on the RHS of (1.11.39), similar to (1.10.22), we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,s,E_n^*(s))\right] \\
=&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s\left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,s,E_n^*(s))\right] + \sum_{s\in\mathcal{S}} r_n(u,\tau,s,E_n^*(s)), \qquad (1.11.40)
\end{aligned}$$

where $\{\xi_i^s\}_{i=1}^n$ is a sequence of i.i.d. Poisson(1) random variables and is independent of everything else, and

$$r_n(u,\tau,s,e) = \text{sign}(N(n_1(s)) - n_1(s))\sum_{j=1}^{\infty}\frac{\#I_n^j(s)}{\sqrt{n}}\frac{1}{\#I_n^j(s)}\sum_{i\in I_n^j(s)}\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right].$$

We aim to show

$$\sup_{|e|\le M,\tau\in\Upsilon,s\in\mathcal{S}}|r_n(u,\tau,s,e)| = o_p(1), \qquad (1.11.41)$$

Recall that the proof of Lemma 1.10.5 relies on (1.10.21) and the fact that

$$\mathbb{E}\sup_{n\ge k\ge n_0}\sup_{\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\eta}_{j,1}(s,\tau)\right| \to 0.$$

Using the same argument and replacing $\tilde{\eta}_{j,1}(s,\tau)$ by $\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right]$, in order to show (1.11.41), we only need to verify that, as $n\to\infty$ followed by $n_0\to\infty$,

$$\mathbb{E}\sup_{n\ge k\ge n_0}\sup_{|e|\le M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right]\right| \to 0$$

Because $\sup_{|e|\le M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right]\right|$ is bounded as shown below, it suffices to show that, for any $\varepsilon>0$, as $n\to\infty$ followed by $n_0\to\infty$,

$$\mathbb{P}\left(\sup_{n\ge k\ge n_0}\sup_{|e|\le M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right]\right| \ge \varepsilon\right) \to 0. \qquad (1.11.42)$$

Define the class of functions $\mathcal{F}_n$ as

$$\mathcal{F}_n = \{\sqrt{n}\left[\phi_i(u,\tau,s,e) - \mathbb{E}\phi_i(u,\tau,s,e)\right] : |e|\le M, \tau\in\Upsilon, s\in\mathcal{S}\}.$$

Then, $\mathcal{F}_n$ is nested by a VC-class with fixed VC-index. In addition, for fixed u, $\mathcal{F}_n$ has a bounded (and independent of n) envelope function

$$F = |u'\iota_1| + M\left(\max_{\tau\in\Upsilon}|q(\tau)| + |u_1|\right).$$

Last, define $\mathcal{I}_l = \{2^l, 2^l+1, \cdots, 2^{l+1}-1\}$. Then,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{n\geq k\geq n_0}\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|\geq\varepsilon\right)\\
\leq&\sum_{l=\lfloor\log_2(n_0)\rfloor}^{\lfloor\log_2(n)\rfloor+1}\mathbb{P}\left(\sup_{k\in\mathcal{I}_l}\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\frac{1}{k}\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|\geq\varepsilon\right)\\
\leq&\sum_{l=\lfloor\log_2(n_0)\rfloor}^{\lfloor\log_2(n)\rfloor+1}\mathbb{P}\left(\sup_{k\leq 2^{l+1}}\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{j=1}^{k}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|\geq\varepsilon 2^l\right)\\
\leq&\sum_{l=\lfloor\log_2(n_0)\rfloor}^{\lfloor\log_2(n)\rfloor+1}9\mathbb{P}\left(\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{j=1}^{2^{l+1}}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|\geq\varepsilon 2^l/30\right)\\
\leq&\sum_{l=\lfloor\log_2(n_0)\rfloor}^{\lfloor\log_2(n)\rfloor+1}\frac{270\mathbb{E}\sup_{|e|\leq M,\tau\in\Upsilon,s\in\mathcal{S}}\left|\sum_{j=1}^{2^{l+1}}\sqrt{n}\left[\phi_i(u,\tau,s,e)-\mathbb{E}\phi_i(u,\tau,s,e)\right]\right|}{\varepsilon 2^l}\\
\leq&\sum_{l=\lfloor\log_2(n_0)\rfloor}^{\lfloor\log_2(n)\rfloor+1}\frac{C_1}{\varepsilon 2^{l/2}}\\
\leq&\frac{2C_1}{\varepsilon\sqrt{n_0}}\rightarrow 0,
\end{aligned}$$

where the first inequality holds by the union bound, the second inequality holds because on $\mathcal{I}_l$, $2^{l+1}\geq k\geq 2^l$, the third inequality follows the same argument in the proof of Theorem 1.3.1, the fourth inequality is due to the Markov inequality, the fifth inequality follows the standard maximal inequality such as van der Vaart and Wellner (1996, Theorem 2.14.1) and the constant C_1 is independent of (l,ε,n), and the last inequality holds by letting $n\rightarrow\infty$. Because ε is arbitrary, we have established (1.11.42), and thus, (1.11.41), which further implies that

$$\sup_{\tau\in\Upsilon,s\in\mathcal{S}}|r_n(u,\tau,s,E_n^*(s))| = o_p(1),$$

For the leading term of (1.11.40), we have

$$
\begin{aligned}
&\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} \xi_i^s \left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,s,E_n^*(s))\right] \\
&= \sum_{s\in\mathcal{S}} \left[\Gamma_n^{s*}(N(s),\tau,E_n^*(s)) - \Gamma_n^{s*}(N(s)+n_1(s),\tau,E_n^*(s))\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma_n^{s*}(k,\tau,e) = &\sum_{i=1}^{k} \xi_i^s \int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1) \le q_1(\tau)+v\} - 1\{Y_i^s(1) \le q_1(\tau)\}\right) dv \\
&- k\mathbb{E}\left[\int_0^{\frac{u'\iota_1 - e(q(\tau)+\frac{u_1}{\sqrt{n}})}{\sqrt{n}}} \left(1\{Y_i^s(1) \le q_1(\tau)+v\} - 1\{Y_i^s(1) \le q_1(\tau)\}\right) dv\right].
\end{aligned}
$$

By the same argument in (1.11.8), we can show that

$$
\sup_{0<t\le 1,\tau\in\Upsilon,|e|\le M} |\Gamma_n^{s*}(k,\tau,e)| = o_p(1),
$$

where we need to use the fact that the Poisson(1) random variable has an exponential tail and thus

$$
\mathbb{E}\sup_{i\in\{1,\cdots,n\},s\in\mathcal{S}} \xi_i^s = O(\log(n)).
$$

Therefore,

$$
\sup_{\tau\in\Upsilon}\left|\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\left[\phi_i(u,\tau,s,E_n^*(s)) - \mathbb{E}\phi_i(u,\tau,E_n^*(s))\right]\right| = o_p(1). \tag{1.11.43}
$$

For the second term on the RHS of (1.11.39), we have

$$
\begin{aligned}
\sum_{s\in\mathcal{S}}\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni}\mathbb{E}\phi_i(u,\tau,s,e) &= \sum_{s\in\mathcal{S}} n_1^*(s)\mathbb{E}\phi_i(u,\tau,s,e) \\
&= \sum_{s\in\mathcal{S}} \pi p(s)\frac{f_1(q_1(\tau)|s)}{2}(u'\iota_1 - eq(\tau))^2 + o(1), \qquad (1.11.44)
\end{aligned}
$$

where the $o(1)$ term holds uniformly over $(\tau,e)\in\Upsilon\times[-M,M]$, the first equality holds because $\sum_{i=N(s)+1}^{N(s)+n_1(s)} M_{ni} = n_1^*(s)$ and the second equality holds by the same calculation in (1.11.10) and the facts that $n^*(s)/n \xrightarrow{p} p(s)$ and

$$\frac{n_1^*(s)}{n} = \frac{D_n^*(s) + \pi n^*(s)}{n} \xrightarrow{p} \pi p(s).$$

Combining (1.11.25), (1.11.39), (1.11.43), (1.11.44), and the facts that $E_n^*(s) = \frac{n}{n^*(s)} \frac{D_n^*(s)}{\sqrt{n}}$ and $\frac{n}{n^*(s)} \xrightarrow{p} 1/p(s)$, we have

$$L_{2,1,n}^*(u,\tau) = \frac{\pi f_1(q_1(\tau))}{2}(u'\iota_1)^2 - \sum_{s\in\mathcal{S}} f_1(q_1(\tau)|s)\frac{\pi D_n^*(s)u'\iota_1}{\sqrt{n}} q(\tau) + h_{2,1}^*(\tau) + R_{sfe,2,1}^*(u,\tau),$$

where

$$h_{2,1}^*(\tau) = \sum_{s\in\mathcal{S}} \frac{\pi f_1(q_1(\tau)|s)}{2} p(s)(E_n^*(s))^2 q^2(\tau)$$

and

$$\sup_{\tau\in\Upsilon} |R_{sfe,2,1}^*(u,\tau)| = o_p(1).$$

This concludes the proof. □

Lemma 1.11.7. *Recall the definition of* $(W_{sfe,n,1}^*(\tau) - \mathcal{W}_{n,1}(\tau), W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau))$ *in* (1.11.27). *If Assumptions in Theorem 1.5.1 hold, then conditionally on data,*

$$(W_{sfe,n,1}^*(\tau) - \mathcal{W}_{n,1}(\tau), W_{sfe,n,2}^*(\tau), W_{sfe,n,3}^*(\tau) - \mathcal{W}_{n,2}(\tau)) \rightsquigarrow (\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau)),$$

where $(\mathcal{B}_1(\tau), \mathcal{B}_2(\tau), \mathcal{B}_3(\tau))$ *are three independent Gaussian processes with covariance kernels*

$$\Sigma_1(\tau_1,\tau_2) = \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_1(S,\tau_1)m_1(S,\tau_2)}{\pi f_1(q_1(\tau_1))f_1(q_1(\tau_2))} + \frac{\min(\tau_1,\tau_2) - \tau_1\tau_2 - \mathbb{E}m_0(S,\tau_1)m_0(S,\tau_2)}{(1-\pi)f_0(q_0(\tau_1))f_0(q_0(\tau_2))},$$

$$\begin{aligned}
&\Sigma_2(\tau_1,\tau_2) \\
=&\mathbb{E}\gamma(S)\Bigg[(m_1(S,\tau_1) - m_0(S,\tau_1))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_1))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_1))}\right) + q(\tau_1)\left(\frac{f_1(q(\tau_1)|S)}{f_1(q_1(\tau_1))} - \frac{f_0(q(\tau_1)|S)}{f_0(q_0(\tau_1))}\right) \\
&\times\Bigg[(m_1(S,\tau_2) - m_0(S,\tau_2))\left(\frac{1-\pi}{\pi f_1(q_1(\tau_2))} - \frac{\pi}{(1-\pi)f_0(q_0(\tau_2))}\right) + q(\tau_2)\left(\frac{f_1(q(\tau_2)|S)}{f_1(q_2(\tau_2))} - \frac{f_0(q(\tau_2)|S)}{f_0(q_0(\tau_2))}\right)\Bigg],
\end{aligned}$$

and

$$\Sigma_3(\tau_1,\tau_2) = \mathbb{E}\left[\frac{m_1(S,\tau_1)}{f_1(q_1(\tau_1))} - \frac{m_0(S,\tau_1)}{f_0(q_0(\tau_1))}\right]\left[\frac{m_1(S,\tau_2)}{f_1(q_1(\tau_2))} - \frac{m_0(S,\tau_2)}{f_0(q_0(\tau_2))}\right],$$

respectively.

Proof. Let $\mathcal{A}_n = \{(A_i^*, S_i^*, A_i, S_i) : i = 1, \cdots, n\}$. Following the definition of M_{ni} and arguments in the proof of Lemma 1.10.5, we have

$$\begin{aligned}
&\{W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau)|\mathcal{A}_n\}\\
&\stackrel{d}{=}\left\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(M_{ni}-1)\left(\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))}\right) - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(M_{ni}-1)\left(\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right)\right]\middle|\mathcal{A}_n\right\}\\
&=\left\{\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] + R_1(\tau)\middle|\mathcal{A}_n\right\},
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R_1(\tau)| = o_p(1)$ and $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$ are sequences of i.i.d. Poisson(1) random variables that are independent of $\mathcal{A}_n$ and across $s \in \mathcal{S}$. In addition, by the same argument in the proof of Lemma 1.10.2, we have

$$\begin{aligned}
&\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=N(s)+1}^{N(s)+n_1(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right]\\
&=\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\left[\sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi p(s))\rfloor}(\xi_i^s-1)\frac{\tilde{\eta}_{i,1}(s,\tau)}{\pi f_1(q_1(\tau))} - \sum_{i=\lfloor n(F(s)+\pi p(s))\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor}(\xi_i^s-1)\frac{\tilde{\eta}_{i,0}(s,\tau)}{(1-\pi)f_0(q_0(\tau))}\right] + R_2(\tau)\\
&\equiv W_1^*(\tau) + R_2(\tau),
\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R_2(\tau)| = o_p(1)$. Because both $W^*_{sfe,n,2}(\tau)$ and $W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)$ are in the σ-field generated by $\mathcal{A}_n$, we have

$$\begin{aligned}
&(W^*_{sfe,n,1}(\tau) - \mathcal{W}_{n,1}(\tau), W^*_{sfe,n,2}(\tau), W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau))\\
\stackrel{d}{=}&(W_1^*(\tau) + R_1(\tau) + R_2(\tau), W^*_{sfe,n,2}(\tau), W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)).
\end{aligned}$$

In addition, note that $\{\xi_i^s\}_{i=1}^n$ and $\{\tilde{\eta}_{i,1}(s,\tau), \tilde{\eta}_{i,1}(s,\tau)\}_{i=1}^n$ are independent of $\mathcal{A}_n$, therefore, $W_1^*(\tau) \perp\!\!\!\perp (W^*_{sfe,n,2}(\tau), W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau))$. Applying van der Vaart and Wellner (1996, Theorem 2.9.6) to each segment

$$\lfloor nF(s)\rfloor + 1, \cdots, \lfloor n(F(s) + \pi p(s))\rfloor \quad \text{or} \quad \lfloor n(F(s) + \pi p(s))\rfloor + 1, \cdots, \lfloor n(F(s) + p(s))\rfloor$$

for $s \in \mathcal{S}$ and noticing that $\{\tilde{\eta}_{i,1}(s,\tau)\}_{i=1}^n$ and $\{\tilde{\eta}_{i,0}(s,\tau)\}_{i=1}^n$ are two i.i.d. sequences for each $s \in \mathcal{S}$, independent of each other, and independent across s, we have, conditionally on $\{\tilde{\eta}_{i,1}(s,\tau), \tilde{\eta}_{i,0}(s,\tau)\}_{i=1}^n$, $s \in \mathcal{S}$,

$$W_1^*(\tau) \rightsquigarrow \mathcal{B}_1(\tau)$$

with the covariance kernel $\Sigma_1(\tau_1, \tau_2)$.

For $W^*_{sfe,n,2}(\tau)$, we note that it depends on data only through $\{S_i^*\}_{i=1}^n$. By Assumption 4,

$$W^*_{sfe,n,2}(\tau)|\{S_i^*\}_{i=1}^n \rightsquigarrow \mathcal{B}_2(\tau)$$

with the covariance kernel $\Sigma_2(\tau_1, \tau_2)$.

Last, for $W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau)$, note that $\{S_i^*\}$ is sampled by the standard bootstrap procedure. Therefore, directly applying van der Vaart and Wellner (1996, Theorem 3.6.2), we have

$$W^*_{sfe,n,3}(\tau) - \mathcal{W}_{n,2}(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\xi_i' - 1)\left[\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right] + R_3(\tau)$$

where $\sup_{\tau\in\Upsilon}|R_3(\tau)| = o_p(1)$, $\{\xi_i'\}_{i=1}^n$ is a sequence of i.i.d. Poisson(1) random variables that is independent of data and $\{\xi_i^s\}_{i=1}^n$, $s \in \mathcal{S}$. By van der Vaart and Wellner (1996, Theorem 3.6.2), conditionally on data $\{S_i\}_{i=1}^n$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (\xi_i' - 1)\left[\frac{m_1(S_i,\tau)}{f_1(q_1(\tau))} - \frac{m_0(S_i,\tau)}{f_0(q_0(\tau))}\right] \rightsquigarrow \mathcal{B}_3(\tau),$$

where $\mathcal{B}_3(\tau)$ has the covariance kernel $\Sigma_3(\tau_1, \tau_2)$. Furthermore, $\mathcal{B}_2(\tau)$ and $\mathcal{B}_3(\tau)$ are independent as $\Sigma_2(\tau_1, \tau_2)$ is not a function of $\{S_i^*\}_{i=1}^n$. This concludes the proof.

□

1.11.6 Additional Simulation Results

1.11.7 DGPs

We consider the following four DGPs with parameters $\gamma = 4$, $\sigma = 2$, and μ which will be specified later. DGPs 1 and 3 correspond to DGPs 1 and 2 in Section 1.6 in the main paper.

1. Let Z be the standardized Beta(2, 2) distributed, $S_i = \sum_{j=1}^{4}\{Z_i \le g_j\}$, and $(g_1, \cdots, g_4) = (-0.25\sqrt{20}, 0, 0.25\sqrt{20}, 0.5\sqrt{20})$. The outcome equation is

$$Y_i = A_i\mu + \gamma Z_i + \eta_i,$$

where $\eta_i = \sigma A_i \varepsilon_{i,1} + (1 - A_i)\varepsilon_{i,2}$ and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

2. Let S be the same as in DGP1. The outcome equation is

$$Y_i = A_i\mu + \gamma Z_i A_i - \gamma(1 - A_i)(\log(Z_i + 3)1\{Z_i \le 0.5\}) + \eta_i.$$

where $\eta_i = \sigma A_i \varepsilon_{i,1} + (1 - A_i)\varepsilon_{i,2}$ and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

3. Let Z be uniformly distributed on $[-2, 2]$, $S_i = \sum_{j=1}^{4}\{Z_i \le g_j\}$, and $(g_1, \cdots, g_4) = (-1, 0, 1, 2)$. The outcome equation is

$$Y_i = A_i\mu + A_i m_{i,1} + (1 - A_i)m_{i,0} + \eta_i,$$

where $m_{i,0} = \gamma Z_i^2 1\{|Z_i| \ge 1\} + \frac{\gamma}{4}(2 - Z_i^2)1\{|Z_i| < 1\}$, $\eta_i = \sigma(1 + Z_i^2)A_i\varepsilon_{i,1} + (1 + Z_i^2)(1 - A_i)\varepsilon_{i,2}$, and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are mutually independent $T(3)/3$ distributed.

4. Let Z_i be normally distributed with mean 0 and variance 4, $S_i = \sum_{j=1}^{4}\{Z_i \le g_j\}$, $(g_1, \cdots, g_4) = (2\Phi^{-1}(0.25), 2\Phi^{-1}(0.5), 2\Phi^{-1}(0.75), \infty)$, and $\Phi(\cdot)$ is the standard normal CDF. The outcome equation is

$$Y_i = A_i\mu + A_i m_{i,1} + (1 - A_i)m_{i,0} + \eta_i,$$

where $m_{i,0} = -\gamma Z_i^2/4$, $m_{i,1} = \gamma Z_i^2/4$,

$$\eta_i = \sigma(1 + 0.5\exp(-Z_i^2/2))A_i\varepsilon_{i,1} + (1 + 0.5\exp(-Z_i^2/2))(1 - A_i)\varepsilon_{i,2},$$

and $(\varepsilon_{i,1}, \varepsilon_{i,2})$ are jointly standard normal.

When $\pi = \frac{1}{2}$, for each DGP, we consider four randomization schemes:

1. SRS: Treatment assignment is generated as in Example 1.

2. WEI: Treatment assignment is generated as in Example 2 with $\phi(x) = (1-x)/2$.

3. BCD: Treatment assignment is generated as in Example 3 with $\lambda = 0.75$.

4. SBR: Treatment assignment is generated as in Example 4.

When $\pi \neq 0.5$, we focus on SRS and SBR. We conduct the simulations with sample sizes $n = 200$ and 400. The numbers of simulation replications and bootstrap samples are 1000. Under the null, $\mu = 0$ and the true parameters of interest are computed by simulations with 10^6 sample size and 10^4 replications. Under the alternative, we perturb the true values by $\mu = 1$ and $\mu = 0.75$ for $n = 200$ and 400, respectively. We consider the following eight t-statistics.

1. "s/naive": the point estimator is computed by the simple QR and its standard error σ_{naive} is computed as

$$\begin{aligned}\sigma^2_{naive} =& \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_1^2(S_i,\tau)}{\pi \hat{f}_1^2(\hat{q}_1(\tau))} + \frac{\tau(1-\tau) - \frac{1}{n}\sum_{i=1}^n \hat{m}_0^2(S_i,\tau)}{(1-\pi)\hat{f}_0^2(\hat{q}_0(\tau))} \\ &+ \frac{1}{n}\sum_{i=1}^n \pi(1-\pi)\left(\frac{\hat{m}_1(S_i,\tau)}{\pi \hat{f}_1(\hat{q}_1(\tau))} + \frac{\hat{m}_0(S_i,\tau)}{(1-\pi)\hat{f}_0(\hat{q}_0(\tau))}\right)^2 \\ &+ \frac{1}{n}\sum_{i=1}^n \left(\frac{\hat{m}_1(S_i,\tau)}{\hat{f}_1(\hat{q}_1(\tau))} - \frac{\hat{m}_0(S_i,\tau)}{\hat{f}_0(\hat{q}_0(\tau))}\right)^2, \end{aligned} \tag{1.11.45}$$

where $\hat{q}_j(\tau)$ is the τ-the empirical quantile of $Y_i|A_i = j$,

$$\hat{m}_{i,1}(s,\tau) = \frac{\sum_{i=1}^n A_i 1\{S_i = s\}(\tau - 1\{Y_i \leq \hat{q}_1(\tau)\})}{n_1(s)},$$

$$\hat{m}_{i,0}(s,\tau) = \frac{\sum_{i=1}^n (1-A_i) 1\{S_i = s\}(\tau - 1\{Y_i \leq \hat{q}_0(\tau)\})}{n(s) - n_1(s)},$$

and for $j = 0, 1$, $\hat{f}_j(\cdot)$ is computed by the kernel density estimation using the observations Y_i provided that $A_i = j$, bandwidth $h_j = 1.06\hat{\sigma}_j n_j^{-1/5}$, and the Gaussian kernel function, where $\hat{\sigma}_j$ is the standard deviation of the observations Y_i provided that $A_i = j$, and $n_j = \sum_{i=1}^n \{A_i = j\}$, $j = 0, 1$.

2. "s/adj": exactly the same as the "s/naive" method with one difference: replacing $\pi(1-\pi)$ in σ^2_{naive} by $\gamma(S_i)$.

3. "s/W": the point estimator is computed by the simple QR and its standard error σ_B is computed by the weighted bootstrap procedure. The bootstrap weights $\{\xi_i\}_{i=1}^n$ are generated from the standard exponential distribution. Denote $\{\hat{\beta}_{1,b}^w\}_{b=1}^B$ as the collection of B estimates obtained by the simple QR applied to the samples generated by the weighted bootstrap procedure. Then,

$$\sigma_B = \frac{\hat{Q}(0.9) - \hat{Q}(0.1)}{\Phi^{-1}(0.9) - \Phi^{-1}(0.1)},$$

where $\Phi(\cdot)$ is the standard normal CDF and $\hat{Q}(\tau)$ is the τ-th empirical quantile of $\{\hat{\beta}_{1,b}^w\}_{b=1}^B$.

4. "sfe/W": the same as above with one difference: the estimation method for both the original and bootstrap samples is the QR with strata fixed effects.

5. "ipw/W": the same as above with one difference: the estimation method for both the original and bootstrap samples is the inverse propensity score weighted QR.

6. "s/CA": the point estimator is computed by the simple QR and its standard error σ_{CA} is computed by the covariate-adaptive bootstrap procedure. Denote $\{\hat{\beta}_{1,b}^*\}_{b=1}^B$ as the collection of B estimates obtained by the simple QR applied to the samples generated by the covariate-adaptive bootstrap procedure. Then,

$$\sigma_{CA} = \frac{\hat{Q}(0.9) - \hat{Q}(0.1)}{\Phi^{-1}(0.9) - \Phi^{-1}(0.1)},$$

where $\hat{Q}(\tau)$ is the τ-th empirical quantile of $\{\hat{\beta}_{1,b}^*\}_{b=1}^B$.

7. "sfe/CA": the same as above with one difference: the estimation method for both the original and bootstrap samples is the QR with strata fixed effects.

8. "ipw/CA": the same as above with one difference: the estimation method for both the original and bootstrap samples is the inverse propensity score weighted QR.

1.11.8 QTE, H_0, $\pi = 0.5$

Table IX. H_0, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.042	0.042	0.051	0.039	0.047	0.046	0.044	0.046
	WEI	0.011	0.038	0.018	0.043	0.046	0.037	0.047	0.047
	BCD	0.004	0.041	0.010	0.043	0.043	0.045	0.048	0.048
	SBR	0.003	0.047	0.003	0.047	0.054	0.049	0.046	0.046
2	SRS	0.045	0.045	0.060	0.062	0.066	0.056	0.069	0.069
	WEI	0.023	0.037	0.049	0.056	0.066	0.068	0.064	0.068
	BCD	0.021	0.037	0.032	0.049	0.057	0.063	0.059	0.057
	SBR	0.025	0.042	0.037	0.050	0.054	0.057	0.054	0.053
3	SRS	0.042	0.042	0.045	0.045	0.054	0.055	0.044	0.058
	WEI	0.042	0.043	0.037	0.044	0.045	0.045	0.043	0.045
	BCD	0.052	0.056	0.044	0.050	0.057	0.057	0.057	0.055
	SBR	0.046	0.053	0.041	0.043	0.048	0.052	0.048	0.047
4	SRS	0.054	0.054	0.048	0.046	0.049	0.046	0.043	0.048
	WEI	0.050	0.051	0.045	0.035	0.047	0.051	0.043	0.055
	BCD	0.056	0.059	0.040	0.030	0.049	0.047	0.044	0.048
	SBR	0.061	0.065	0.044	0.032	0.053	0.057	0.051	0.053

Table X. H_0, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.045	0.045	0.047	0.043	0.044	0.044	0.039	0.039
	WEI	0.012	0.040	0.014	0.044	0.043	0.037	0.041	0.035
	BCD	0.002	0.057	0.003	0.040	0.041	0.044	0.039	0.039
	SBR	0.001	0.057	0.001	0.045	0.046	0.045	0.045	0.044
2	SRS	0.045	0.045	0.057	0.066	0.061	0.048	0.064	0.066
	WEI	0.033	0.065	0.037	0.056	0.065	0.065	0.056	0.061
	BCD	0.022	0.062	0.027	0.048	0.056	0.057	0.057	0.054
	SBR	0.017	0.050	0.017	0.040	0.046	0.048	0.048	0.046
3	SRS	0.004	0.004	0.047	0.045	0.052	0.052	0.047	0.053
	WEI	0.006	0.006	0.045	0.050	0.058	0.052	0.053	0.057
	BCD	0.010	0.010	0.045	0.050	0.051	0.050	0.050	0.053
	SBR	0.008	0.011	0.048	0.048	0.053	0.046	0.051	0.047
4	SRS	0.013	0.013	0.050	0.036	0.051	0.055	0.035	0.043
	WEI	0.011	0.011	0.043	0.033	0.051	0.049	0.043	0.052
	BCD	0.013	0.013	0.049	0.041	0.053	0.055	0.047	0.052
	SBR	0.013	0.013	0.040	0.033	0.047	0.046	0.044	0.045

Table XI. H_0, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.052	0.052	0.053	0.044	0.044	0.048	0.041	0.042
	WEI	0.012	0.042	0.014	0.043	0.046	0.037	0.039	0.045
	BCD	0.002	0.047	0.002	0.051	0.054	0.055	0.053	0.053
	SBR	0.001	0.026	0.003	0.030	0.035	0.030	0.033	0.035
2	SRS	0.052	0.052	0.066	0.057	0.058	0.053	0.048	0.058
	WEI	0.021	0.045	0.027	0.047	0.052	0.057	0.051	0.054
	BCD	0.013	0.046	0.025	0.051	0.060	0.067	0.061	0.060
	SBR	0.008	0.036	0.012	0.037	0.046	0.046	0.046	0.050
3	SRS	0.058	0.058	0.048	0.054	0.047	0.058	0.054	0.051
	WEI	0.053	0.055	0.041	0.044	0.047	0.047	0.048	0.046
	BCD	0.042	0.043	0.026	0.026	0.033	0.033	0.032	0.034
	SBR	0.048	0.052	0.040	0.036	0.046	0.051	0.043	0.048
4	SRS	0.044	0.044	0.057	0.059	0.062	0.053	0.051	0.065
	WEI	0.034	0.034	0.044	0.029	0.053	0.048	0.044	0.054
	BCD	0.029	0.032	0.040	0.019	0.045	0.047	0.043	0.047
	SBR	0.034	0.037	0.042	0.025	0.051	0.055	0.049	0.051

Table XII. H_0, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.047	0.047	0.053	0.041	0.039	0.049	0.040	0.040
	WEI	0.009	0.043	0.017	0.041	0.042	0.045	0.044	0.043
	BCD	0.002	0.042	0.003	0.037	0.040	0.035	0.036	0.037
	SBR	0.002	0.043	0.004	0.034	0.034	0.036	0.032	0.030
2	SRS	0.046	0.046	0.056	0.059	0.059	0.055	0.057	0.059
	WEI	0.035	0.046	0.046	0.056	0.062	0.065	0.061	0.060
	BCD	0.030	0.044	0.037	0.055	0.065	0.060	0.060	0.057
	SBR	0.026	0.049	0.042	0.058	0.067	0.063	0.062	0.066
3	SRS	0.044	0.044	0.039	0.041	0.042	0.042	0.041	0.043
	WEI	0.042	0.045	0.048	0.041	0.048	0.051	0.046	0.049
	BCD	0.039	0.040	0.041	0.040	0.044	0.046	0.047	0.048
	SBR	0.048	0.051	0.046	0.048	0.052	0.056	0.056	0.055
4	SRS	0.056	0.056	0.039	0.042	0.041	0.041	0.043	0.042
	WEI	0.052	0.055	0.038	0.034	0.045	0.042	0.044	0.044
	BCD	0.054	0.058	0.040	0.026	0.045	0.044	0.045	0.043
	SBR	0.061	0.068	0.049	0.027	0.047	0.054	0.055	0.051

Table XIII. H_0, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.042	0.042	0.054	0.046	0.040	0.046	0.050	0.041
	WEI	0.010	0.049	0.008	0.047	0.047	0.046	0.043	0.042
	BCD	0.003	0.045	0.002	0.043	0.043	0.035	0.039	0.040
	SBR	0.002	0.046	0.000	0.035	0.037	0.036	0.036	0.037
2	SRS	0.050	0.050	0.055	0.049	0.047	0.051	0.052	0.050
	WEI	0.018	0.048	0.025	0.041	0.046	0.045	0.048	0.045
	BCD	0.011	0.042	0.011	0.041	0.046	0.045	0.046	0.043
	SBR	0.017	0.051	0.014	0.042	0.050	0.053	0.047	0.050
3	SRS	0.012	0.012	0.043	0.046	0.048	0.046	0.050	0.050
	WEI	0.014	0.016	0.057	0.055	0.060	0.055	0.058	0.057
	BCD	0.013	0.013	0.055	0.059	0.061	0.051	0.053	0.052
	SBR	0.006	0.006	0.040	0.040	0.039	0.038	0.039	0.038
4	SRS	0.019	0.019	0.056	0.052	0.064	0.056	0.051	0.061
	WEI	0.018	0.018	0.060	0.046	0.065	0.064	0.062	0.066
	BCD	0.015	0.015	0.057	0.046	0.066	0.063	0.059	0.067
	SBR	0.021	0.021	0.057	0.043	0.060	0.062	0.062	0.062

Table XIV. H_0, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.051	0.051	0.056	0.055	0.056	0.052	0.055	0.054
	WEI	0.007	0.041	0.014	0.055	0.053	0.051	0.050	0.051
	BCD	0.006	0.038	0.004	0.046	0.048	0.041	0.042	0.046
	SBR	0.004	0.033	0.002	0.044	0.043	0.042	0.043	0.042
2	SRS	0.048	0.048	0.073	0.055	0.061	0.060	0.057	0.059
	WEI	0.020	0.039	0.024	0.046	0.053	0.048	0.051	0.053
	BCD	0.012	0.048	0.020	0.050	0.051	0.057	0.055	0.051
	SBR	0.011	0.047	0.014	0.046	0.052	0.050	0.052	0.052
3	SRS	0.054	0.054	0.050	0.045	0.052	0.049	0.044	0.052
	WEI	0.053	0.055	0.049	0.047	0.053	0.050	0.049	0.054
	BCD	0.059	0.063	0.038	0.041	0.045	0.044	0.043	0.043
	SBR	0.049	0.051	0.042	0.044	0.043	0.049	0.049	0.049
4	SRS	0.054	0.054	0.057	0.053	0.063	0.055	0.056	0.063
	WEI	0.047	0.051	0.055	0.043	0.064	0.055	0.061	0.059
	BCD	0.049	0.051	0.054	0.033	0.063	0.062	0.056	0.063
	SBR	0.046	0.048	0.047	0.026	0.051	0.057	0.056	0.053

1.11.9 QTE, H_1, $\pi = 0.5$

Table XV. H_1, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.191	0.191	0.203	0.354	0.356	0.205	0.340	0.342
	WEI	0.126	0.257	0.147	0.359	0.358	0.279	0.345	0.350
	BCD	0.105	0.372	0.122	0.379	0.375	0.361	0.369	0.365
	SBR	0.099	0.400	0.114	0.378	0.382	0.411	0.375	0.368
2	SRS	0.284	0.284	0.315	0.352	0.376	0.319	0.345	0.378
	WEI	0.270	0.319	0.314	0.356	0.364	0.359	0.363	0.369
	BCD	0.282	0.333	0.304	0.361	0.375	0.390	0.385	0.383
	SBR	0.290	0.346	0.296	0.335	0.361	0.387	0.358	0.356
3	SRS	0.712	0.712	0.694	0.688	0.698	0.704	0.677	0.686
	WEI	0.701	0.707	0.678	0.685	0.680	0.699	0.687	0.674
	BCD	0.712	0.720	0.673	0.686	0.695	0.699	0.698	0.698
	SBR	0.672	0.684	0.659	0.639	0.647	0.673	0.647	0.638
4	SRS	0.166	0.166	0.124	0.112	0.132	0.135	0.131	0.128
	WEI	0.166	0.170	0.126	0.098	0.125	0.144	0.139	0.133
	BCD	0.165	0.176	0.126	0.094	0.155	0.157	0.145	0.157
	SBR	0.167	0.175	0.122	0.088	0.139	0.145	0.133	0.140

Table XVI. H_1, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.183	0.183	0.193	0.443	0.441	0.200	0.431	0.429
	WEI	0.116	0.295	0.138	0.442	0.447	0.298	0.437	0.436
	BCD	0.072	0.472	0.095	0.450	0.453	0.434	0.446	0.448
	SBR	0.085	0.485	0.099	0.463	0.460	0.457	0.453	0.448
2	SRS	0.267	0.267	0.256	0.359	0.366	0.265	0.358	0.371
	WEI	0.248	0.346	0.247	0.358	0.394	0.346	0.378	0.389
	BCD	0.229	0.402	0.233	0.358	0.396	0.388	0.395	0.392
	SBR	0.232	0.404	0.234	0.365	0.392	0.399	0.401	0.391
3	SRS	0.797	0.797	0.904	0.897	0.916	0.902	0.897	0.913
	WEI	0.802	0.807	0.907	0.903	0.909	0.913	0.902	0.906
	BCD	0.796	0.804	0.902	0.910	0.911	0.908	0.911	0.906
	SBR	0.771	0.774	0.897	0.896	0.901	0.899	0.894	0.899
4	SRS	0.176	0.176	0.312	0.269	0.317	0.316	0.297	0.316
	WEI	0.171	0.175	0.289	0.255	0.307	0.309	0.297	0.298
	BCD	0.169	0.174	0.299	0.262	0.313	0.329	0.311	0.316
	SBR	0.163	0.165	0.283	0.255	0.304	0.302	0.298	0.298

Table XVII. H_1, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.198	0.198	0.215	0.362	0.358	0.216	0.353	0.355
	WEI	0.143	0.293	0.153	0.361	0.368	0.315	0.362	0.364
	BCD	0.108	0.377	0.131	0.356	0.360	0.355	0.353	0.353
	SBR	0.079	0.386	0.105	0.397	0.396	0.381	0.403	0.386
2	SRS	0.268	0.268	0.315	0.386	0.439	0.322	0.391	0.434
	WEI	0.238	0.339	0.285	0.396	0.430	0.390	0.417	0.428
	BCD	0.209	0.407	0.263	0.398	0.428	0.425	0.428	0.418
	SBR	0.206	0.427	0.267	0.439	0.455	0.450	0.465	0.456
3	SRS	0.698	0.698	0.607	0.594	0.619	0.634	0.609	0.622
	WEI	0.668	0.673	0.607	0.606	0.616	0.631	0.623	0.624
	BCD	0.690	0.698	0.607	0.612	0.616	0.635	0.618	0.621
	SBR	0.669	0.675	0.596	0.614	0.633	0.617	0.631	0.630
4	SRS	0.163	0.163	0.158	0.122	0.167	0.173	0.140	0.169
	WEI	0.144	0.152	0.152	0.105	0.175	0.169	0.152	0.178
	BCD	0.133	0.138	0.151	0.085	0.170	0.177	0.173	0.172
	SBR	0.146	0.154	0.143	0.090	0.175	0.171	0.177	0.180

Table XVIII. H_1, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.206	0.206	0.229	0.403	0.417	0.231	0.401	0.405
	WEI	0.163	0.332	0.173	0.408	0.413	0.337	0.408	0.413
	BCD	0.121	0.430	0.143	0.420	0.422	0.421	0.419	0.413
	SBR	0.128	0.451	0.144	0.428	0.429	0.458	0.426	0.423
2	SRS	0.312	0.312	0.345	0.422	0.415	0.351	0.416	0.416
	WEI	0.312	0.352	0.332	0.405	0.424	0.378	0.408	0.426
	BCD	0.299	0.378	0.333	0.392	0.405	0.403	0.415	0.413
	SBR	0.330	0.389	0.345	0.401	0.407	0.426	0.410	0.406
3	SRS	0.763	0.763	0.734	0.730	0.740	0.738	0.732	0.738
	WEI	0.763	0.764	0.739	0.739	0.748	0.744	0.746	0.746
	BCD	0.781	0.783	0.760	0.760	0.768	0.772	0.774	0.767
	SBR	0.766	0.773	0.745	0.739	0.744	0.763	0.751	0.744
4	SRS	0.177	0.177	0.129	0.108	0.136	0.127	0.121	0.133
	WEI	0.170	0.176	0.129	0.096	0.139	0.139	0.131	0.143
	BCD	0.178	0.185	0.132	0.089	0.141	0.141	0.139	0.138
	SBR	0.180	0.186	0.129	0.102	0.134	0.147	0.135	0.133

Table XIX. H_1, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.218	0.218	0.232	0.504	0.502	0.235	0.497	0.502
	WEI	0.147	0.356	0.160	0.503	0.503	0.350	0.498	0.507
	BCD	0.089	0.526	0.117	0.498	0.502	0.493	0.495	0.496
	SBR	0.089	0.550	0.109	0.520	0.518	0.524	0.526	0.519
2	SRS	0.301	0.301	0.309	0.402	0.426	0.306	0.413	0.423
	WEI	0.287	0.387	0.281	0.402	0.418	0.372	0.411	0.420
	BCD	0.268	0.451	0.262	0.400	0.443	0.434	0.434	0.441
	SBR	0.260	0.433	0.252	0.403	0.421	0.418	0.431	0.420
3	SRS	0.897	0.897	0.956	0.957	0.956	0.957	0.956	0.957
	WEI	0.892	0.892	0.954	0.944	0.948	0.951	0.942	0.948
	BCD	0.887	0.889	0.952	0.949	0.954	0.957	0.954	0.956
	SBR	0.900	0.902	0.954	0.954	0.954	0.958	0.962	0.957
4	SRS	0.234	0.234	0.345	0.317	0.351	0.353	0.339	0.343
	WEI	0.222	0.224	0.336	0.326	0.352	0.352	0.335	0.358
	BCD	0.226	0.230	0.346	0.321	0.349	0.368	0.359	0.365
	SBR	0.238	0.242	0.369	0.350	0.380	0.379	0.374	0.377

Table XX. H_1, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.218	0.218	0.237	0.430	0.435	0.242	0.438	0.435
	WEI	0.163	0.321	0.176	0.441	0.437	0.344	0.433	0.432
	BCD	0.136	0.422	0.152	0.421	0.420	0.417	0.417	0.416
	SBR	0.103	0.446	0.124	0.459	0.459	0.448	0.463	0.461
2	SRS	0.300	0.300	0.337	0.445	0.479	0.335	0.449	0.479
	WEI	0.258	0.369	0.313	0.446	0.465	0.414	0.453	0.463
	BCD	0.247	0.462	0.295	0.451	0.476	0.483	0.481	0.477
	SBR	0.227	0.444	0.276	0.472	0.490	0.471	0.496	0.492
3	SRS	0.763	0.763	0.710	0.702	0.707	0.712	0.701	0.715
	WEI	0.773	0.776	0.696	0.701	0.700	0.720	0.709	0.706
	BCD	0.753	0.755	0.705	0.716	0.720	0.720	0.717	0.726
	SBR	0.746	0.750	0.684	0.699	0.705	0.692	0.709	0.708
4	SRS	0.209	0.209	0.199	0.140	0.221	0.208	0.149	0.221
	WEI	0.201	0.208	0.191	0.110	0.203	0.206	0.178	0.204
	BCD	0.195	0.200	0.199	0.121	0.213	0.224	0.213	0.220
	SBR	0.198	0.203	0.198	0.114	0.229	0.214	0.230	0.225

1.11.10 QTE, H_0, $\pi = 0.7$

Table XXI. H_0, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.042	0.042	0.046	0.042	0.036	0.036	0.039	0.039
	SBR	0.002	0.014	0.005	0.053	0.052	0.049	0.050	0.047
2	SRS	0.037	0.037	0.051	0.059	0.057	0.061	0.057	0.064
	SBR	0.032	0.036	0.042	0.046	0.048	0.055	0.055	0.055
3	SRS	0.046	0.046	0.046	0.047	0.039	0.045	0.049	0.043
	SBR	0.040	0.044	0.032	0.031	0.034	0.041	0.037	0.040
4	SRS	0.098	0.098	0.067	0.075	0.069	0.062	0.057	0.066
	SBR	0.057	0.066	0.043	0.016	0.062	0.061	0.066	0.064

Table XXII. H_0, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.048	0.048	0.052	0.045	0.047	0.034	0.040	0.044
	SBR	0.001	0.007	0.002	0.039	0.040	0.044	0.038	0.037
2	SRS	0.057	0.057	0.065	0.051	0.058	0.050	0.051	0.053
	SBR	0.022	0.034	0.021	0.053	0.053	0.050	0.059	0.053
3	SRS	0.016	0.016	0.052	0.046	0.054	0.051	0.048	0.053
	SBR	0.004	0.005	0.039	0.038	0.048	0.045	0.046	0.048
4	SRS	0.009	0.009	0.046	0.037	0.049	0.046	0.045	0.051
	SBR	0.004	0.005	0.036	0.016	0.052	0.049	0.043	0.046

Table XXIII. H_0, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.052	0.052	0.057	0.045	0.049	0.044	0.040	0.043
	SBR	0.002	0.008	0.004	0.033	0.034	0.036	0.036	0.036
2	SRS	0.042	0.042	0.061	0.055	0.067	0.047	0.055	0.068
	SBR	0.006	0.014	0.009	0.029	0.037	0.042	0.039	0.040
3	SRS	0.056	0.056	0.043	0.038	0.054	0.048	0.046	0.054
	SBR	0.055	0.057	0.048	0.042	0.050	0.053	0.052	0.052
4	SRS	0.019	0.019	0.038	0.032	0.046	0.045	0.042	0.042
	SBR	0.022	0.022	0.044	0.028	0.045	0.044	0.038	0.042

Table XXIV. H_0, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.044	0.044	0.054	0.039	0.041	0.038	0.040	0.042
	SBR	0.003	0.015	0.003	0.051	0.052	0.043	0.046	0.046
2	SRS	0.034	0.034	0.057	0.058	0.054	0.062	0.058	0.053
	SBR	0.031	0.034	0.040	0.044	0.049	0.051	0.051	0.051
3	SRS	0.037	0.037	0.029	0.034	0.036	0.033	0.033	0.039
	SBR	0.045	0.049	0.037	0.037	0.042	0.044	0.040	0.041
4	SRS	0.073	0.073	0.044	0.054	0.046	0.045	0.048	0.041
	SBR	0.065	0.076	0.036	0.014	0.060	0.058	0.062	0.060

Table XXV. H_0, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.044	0.044	0.051	0.037	0.039	0.048	0.036	0.037
	SBR	0.001	0.002	0.000	0.035	0.039	0.035	0.040	0.040
2	SRS	0.062	0.062	0.062	0.049	0.049	0.059	0.041	0.048
	SBR	0.015	0.029	0.015	0.034	0.040	0.040	0.042	0.037
3	SRS	0.007	0.007	0.039	0.036	0.042	0.042	0.042	0.047
	SBR	0.006	0.006	0.035	0.037	0.036	0.037	0.041	0.037
4	SRS	0.013	0.013	0.046	0.029	0.061	0.053	0.035	0.054
	SBR	0.009	0.010	0.033	0.025	0.056	0.054	0.052	0.050

Table XXVI. H_0, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.049	0.049	0.053	0.046	0.050	0.043	0.048	0.050
	SBR	0.001	0.006	0.002	0.038	0.041	0.037	0.036	0.036
2	SRS	0.050	0.050	0.065	0.050	0.049	0.056	0.052	0.052
	SBR	0.010	0.019	0.015	0.041	0.048	0.042	0.041	0.041
3	SRS	0.044	0.044	0.031	0.042	0.039	0.032	0.038	0.039
	SBR	0.057	0.059	0.040	0.036	0.044	0.043	0.043	0.043
4	SRS	0.034	0.034	0.051	0.046	0.049	0.051	0.046	0.051
	SBR	0.028	0.028	0.044	0.040	0.045	0.045	0.045	0.046

1.11.11 QTE, H_1, $\pi = 0.7$

Table XXVII. H_1, $n = 200$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.152	0.152	0.176	0.359	0.313	0.187	0.343	0.339
	SBR	0.065	0.186	0.100	0.346	0.336	0.357	0.341	0.338
2	SRS	0.314	0.314	0.334	0.361	0.325	0.347	0.367	0.365
	SBR	0.309	0.334	0.336	0.355	0.368	0.383	0.375	0.376
3	SRS	0.704	0.704	0.671	0.665	0.626	0.685	0.663	0.691
	SBR	0.697	0.716	0.663	0.671	0.669	0.702	0.686	0.688
4	SRS	0.136	0.136	0.097	0.094	0.129	0.106	0.093	0.122
	SBR	0.116	0.127	0.081	0.050	0.103	0.107	0.105	0.106

Table XXVIII. H_1, $n = 200$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.170	0.170	0.172	0.411	0.425	0.167	0.407	0.406
	SBR	0.043	0.212	0.060	0.445	0.455	0.457	0.435	0.434
2	SRS	0.287	0.287	0.280	0.371	0.364	0.275	0.374	0.360
	SBR	0.258	0.327	0.236	0.367	0.387	0.372	0.383	0.381
3	SRS	0.771	0.771	0.891	0.882	0.903	0.895	0.883	0.894
	SBR	0.760	0.769	0.892	0.896	0.911	0.901	0.904	0.900
4	SRS	0.145	0.145	0.265	0.218	0.305	0.264	0.241	0.301
	SBR	0.128	0.136	0.235	0.177	0.288	0.290	0.284	0.287

Table XXIX. H_1, $n = 200$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.181	0.181	0.183	0.342	0.340	0.188	0.340	0.338
	SBR	0.072	0.175	0.076	0.353	0.364	0.342	0.357	0.357
2	SRS	0.279	0.279	0.321	0.404	0.427	0.341	0.400	0.427
	SBR	0.243	0.341	0.293	0.430	0.451	0.430	0.454	0.435
3	SRS	0.662	0.662	0.586	0.559	0.599	0.605	0.569	0.592
	SBR	0.631	0.639	0.572	0.564	0.597	0.594	0.601	0.598
4	SRS	0.150	0.150	0.201	0.164	0.199	0.208	0.189	0.211
	SBR	0.143	0.145	0.193	0.166	0.206	0.206	0.208	0.205

Table XXX. H_1, $n = 400$, $\tau = 0.25$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.181	0.181	0.192	0.351	0.354	0.202	0.346	0.351
	SBR	0.083	0.233	0.113	0.392	0.392	0.407	0.394	0.392
2	SRS	0.362	0.362	0.406	0.403	0.415	0.408	0.415	0.424
	SBR	0.350	0.381	0.388	0.412	0.426	0.426	0.422	0.419
3	SRS	0.781	0.781	0.743	0.751	0.758	0.746	0.750	0.759
	SBR	0.791	0.797	0.752	0.765	0.777	0.781	0.778	0.779
4	SRS	0.160	0.160	0.082	0.072	0.112	0.097	0.095	0.116
	SBR	0.133	0.154	0.091	0.044	0.119	0.119	0.121	0.120

Table XXXI. H_1, $n = 400$, $\tau = 0.5$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.184	0.184	0.187	0.468	0.479	0.194	0.460	0.466
	SBR	0.042	0.220	0.059	0.486	0.498	0.505	0.480	0.482
2	SRS	0.322	0.322	0.298	0.405	0.404	0.303	0.412	0.400
	SBR	0.262	0.342	0.237	0.376	0.399	0.385	0.389	0.389
3	SRS	0.867	0.867	0.939	0.930	0.933	0.941	0.932	0.936
	SBR	0.883	0.888	0.948	0.952	0.952	0.955	0.952	0.952
4	SRS	0.209	0.209	0.327	0.275	0.354	0.341	0.308	0.351
	SBR	0.194	0.217	0.310	0.256	0.365	0.364	0.359	0.356

Table XXXII. H_1, $n = 400$, $\tau = 0.75$

M	A	s/naive	s/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.217	0.217	0.224	0.411	0.409	0.219	0.411	0.408
	SBR	0.103	0.246	0.107	0.419	0.418	0.400	0.421	0.420
2	SRS	0.335	0.335	0.378	0.485	0.505	0.384	0.468	0.501
	SBR	0.278	0.384	0.329	0.479	0.500	0.487	0.504	0.493
3	SRS	0.708	0.708	0.661	0.628	0.665	0.665	0.629	0.672
	SBR	0.705	0.706	0.652	0.631	0.665	0.673	0.672	0.673
4	SRS	0.205	0.205	0.226	0.221	0.245	0.234	0.234	0.240
	SBR	0.205	0.205	0.249	0.209	0.248	0.258	0.256	0.258

1.11.12 ATE, $\pi = 0.5$

Table XXXIII. H_0, $n = 200$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.059	0.057	0.051	0.061	0.055	0.057	0.053	0.048	0.049
	WEI	0.006	0.048	0.062	0.004	0.068	0.068	0.051	0.065	0.065
	BCD	0.001	0.089	0.056	0.000	0.058	0.058	0.071	0.056	0.056
	SBR	0.000	0.067	0.061	0.000	0.064	0.064	0.059	0.061	0.061
2	SRS	0.062	0.061	0.061	0.061	0.059	0.062	0.060	0.057	0.059
	WEI	0.027	0.060	0.050	0.029	0.046	0.054	0.057	0.052	0.053
	BCD	0.014	0.058	0.053	0.016	0.053	0.052	0.052	0.052	0.049
	SBR	0.006	0.045	0.044	0.006	0.045	0.045	0.045	0.045	0.045
3	SRS	0.057	0.056	0.068	0.055	0.061	0.061	0.056	0.064	0.065
	WEI	0.049	0.050	0.057	0.052	0.057	0.056	0.048	0.053	0.053
	BCD	0.057	0.058	0.057	0.057	0.063	0.063	0.057	0.056	0.057
	SBR	0.055	0.058	0.056	0.057	0.060	0.061	0.055	0.055	0.055
4	SRS	0.066	0.067	0.077	0.068	0.069	0.063	0.063	0.070	0.063
	WEI	0.065	0.067	0.070	0.066	0.067	0.068	0.069	0.067	0.070
	BCD	0.068	0.068	0.067	0.065	0.061	0.068	0.065	0.065	0.065
	SBR	0.055	0.055	0.055	0.057	0.057	0.058	0.057	0.057	0.057

Table XXXIV. H_1, $n = 200$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.387	0.385	0.948	0.391	0.946	0.946	0.386	0.944	0.942
	WEI	0.330	0.680	0.944	0.334	0.941	0.940	0.691	0.942	0.941
	BCD	0.275	0.917	0.940	0.272	0.943	0.943	0.884	0.942	0.942
	SBR	0.280	0.942	0.951	0.285	0.950	0.950	0.937	0.945	0.945
2	SRS	0.533	0.532	0.750	0.538	0.746	0.758	0.541	0.746	0.753
	WEI	0.532	0.668	0.748	0.533	0.742	0.750	0.675	0.743	0.749
	BCD	0.541	0.748	0.752	0.544	0.751	0.755	0.733	0.751	0.752
	SBR	0.544	0.774	0.779	0.551	0.772	0.781	0.769	0.775	0.775
3	SRS	0.770	0.769	0.767	0.773	0.768	0.775	0.769	0.754	0.760
	WEI	0.760	0.766	0.763	0.759	0.759	0.768	0.765	0.763	0.761
	BCD	0.767	0.772	0.769	0.762	0.771	0.769	0.772	0.765	0.765
	SBR	0.757	0.762	0.761	0.758	0.770	0.767	0.761	0.764	0.764
4	SRS	0.181	0.182	0.181	0.182	0.171	0.184	0.181	0.180	0.186
	WEI	0.180	0.183	0.182	0.184	0.180	0.184	0.184	0.178	0.179
	BCD	0.170	0.175	0.174	0.177	0.177	0.181	0.182	0.183	0.182
	SBR	0.177	0.178	0.179	0.184	0.180	0.186	0.179	0.178	0.178

Table XXXV. H_0, $n = 400$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.063	0.061	0.042	0.063	0.043	0.045	0.055	0.042	0.042
	WEI	0.005	0.050	0.050	0.006	0.052	0.052	0.052	0.050	0.050
	BCD	0.000	0.067	0.052	0.000	0.059	0.059	0.051	0.059	0.059
	SBR	0.000	0.059	0.058	0.000	0.057	0.057	0.063	0.060	0.060
2	SRS	0.061	0.057	0.055	0.058	0.055	0.054	0.061	0.054	0.051
	WEI	0.018	0.051	0.064	0.019	0.063	0.064	0.052	0.064	0.064
	BCD	0.009	0.045	0.046	0.006	0.046	0.047	0.043	0.049	0.049
	SBR	0.014	0.062	0.060	0.016	0.065	0.065	0.063	0.063	0.063
3	SRS	0.050	0.049	0.050	0.050	0.049	0.051	0.052	0.048	0.048
	WEI	0.046	0.047	0.049	0.047	0.046	0.047	0.048	0.047	0.046
	BCD	0.049	0.049	0.049	0.049	0.050	0.050	0.050	0.050	0.050
	SBR	0.055	0.056	0.056	0.059	0.058	0.059	0.055	0.056	0.056
4	SRS	0.057	0.057	0.055	0.056	0.056	0.059	0.054	0.051	0.056
	WEI	0.051	0.051	0.053	0.052	0.054	0.054	0.051	0.051	0.052
	BCD	0.056	0.056	0.056	0.054	0.056	0.056	0.054	0.053	0.053
	SBR	0.056	0.058	0.058	0.055	0.056	0.057	0.057	0.057	0.057

Table XXXVI. H_1, $n = 400$, $\pi = 0.5$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.422	0.422	0.964	0.416	0.968	0.966	0.415	0.964	0.962
	WEI	0.387	0.732	0.969	0.393	0.969	0.969	0.732	0.967	0.968
	BCD	0.341	0.962	0.971	0.350	0.969	0.968	0.955	0.968	0.968
	SBR	0.357	0.967	0.967	0.368	0.966	0.966	0.967	0.965	0.965
2	SRS	0.572	0.568	0.806	0.579	0.795	0.805	0.568	0.796	0.805
	WEI	0.577	0.723	0.813	0.575	0.814	0.810	0.728	0.811	0.808
	BCD	0.606	0.809	0.813	0.618	0.817	0.821	0.802	0.810	0.810
	SBR	0.601	0.828	0.829	0.603	0.832	0.836	0.830	0.834	0.834
3	SRS	0.804	0.801	0.803	0.798	0.798	0.799	0.804	0.803	0.803
	WEI	0.804	0.804	0.806	0.802	0.800	0.803	0.803	0.803	0.803
	BCD	0.816	0.818	0.820	0.822	0.825	0.825	0.819	0.819	0.819
	SBR	0.821	0.823	0.823	0.816	0.820	0.819	0.822	0.822	0.822
4	SRS	0.228	0.230	0.229	0.225	0.227	0.228	0.234	0.226	0.226
	WEI	0.229	0.230	0.230	0.225	0.223	0.228	0.233	0.235	0.234
	BCD	0.221	0.224	0.225	0.227	0.225	0.231	0.231	0.231	0.233
	SBR	0.224	0.226	0.225	0.224	0.225	0.230	0.235	0.235	0.235

1.11.13 ATE, $\pi = 0.7$

Table XXXVII. H_0, $n = 200$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.050	0.045	0.056	0.051	0.056	0.062	0.046	0.054	0.055
	SBR	0.000	0.004	0.051	0.000	0.061	0.064	0.064	0.060	0.059
2	SRS	0.048	0.055	0.074	0.055	0.049	0.056	0.045	0.049	0.057
	SBR	0.013	0.030	0.041	0.013	0.024	0.051	0.056	0.049	0.051
3	SRS	0.059	0.060	0.066	0.060	0.060	0.064	0.058	0.055	0.064
	SBR	0.051	0.053	0.052	0.053	0.045	0.057	0.056	0.056	0.055
4	SRS	0.057	0.057	0.056	0.058	0.056	0.068	0.054	0.057	0.058
	SBR	0.047	0.050	0.044	0.051	0.037	0.054	0.054	0.055	0.055

Table XXXVIII. H_1, $n = 200$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.329	0.328	0.934	0.336	0.943	0.946	0.326	0.941	0.941
	SBR	0.220	0.631	0.938	0.233	0.946	0.949	0.932	0.943	0.943
2	SRS	0.581	0.578	0.687	0.582	0.619	0.756	0.571	0.601	0.758
	SBR	0.598	0.699	0.747	0.599	0.686	0.768	0.752	0.766	0.764
3	SRS	0.773	0.779	0.758	0.769	0.741	0.784	0.773	0.729	0.782
	SBR	0.771	0.773	0.772	0.777	0.763	0.782	0.782	0.780	0.781
4	SRS	0.149	0.154	0.121	0.153	0.140	0.168	0.154	0.141	0.165
	SBR	0.144	0.151	0.129	0.153	0.118	0.175	0.172	0.170	0.169

Table XXXIX. H_0, $n = 400$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.062	0.059	0.065	0.061	0.056	0.056	0.062	0.060	0.061
	SBR	0.000	0.000	0.034	0.000	0.039	0.040	0.045	0.045	0.044
2	SRS	0.052	0.050	0.087	0.054	0.055	0.052	0.050	0.057	0.051
	SBR	0.013	0.029	0.040	0.012	0.027	0.044	0.042	0.044	0.042
3	SRS	0.042	0.041	0.049	0.045	0.043	0.052	0.040	0.040	0.046
	SBR	0.028	0.028	0.031	0.029	0.025	0.032	0.035	0.036	0.034
4	SRS	0.053	0.055	0.043	0.058	0.053	0.058	0.055	0.050	0.056
	SBR	0.050	0.051	0.043	0.051	0.035	0.054	0.055	0.055	0.053

Table XL. H_1, $n = 400$, $\pi = 0.7$

M	A	s/naive	s/adj	sfe/adj	s/W	sfe/W	ipw/W	s/CA	sfe/CA	ipw/CA
1	SRS	0.384	0.380	0.972	0.381	0.971	0.976	0.382	0.970	0.973
	SBR	0.250	0.736	0.970	0.254	0.972	0.972	0.967	0.973	0.974
2	SRS	0.616	0.628	0.753	0.622	0.693	0.796	0.617	0.690	0.795
	SBR	0.659	0.759	0.806	0.665	0.740	0.827	0.817	0.827	0.827
3	SRS	0.818	0.817	0.805	0.812	0.793	0.821	0.816	0.793	0.829
	SBR	0.833	0.838	0.836	0.831	0.824	0.840	0.838	0.839	0.837
4	SRS	0.177	0.172	0.145	0.180	0.162	0.195	0.181	0.171	0.186
	SBR	0.181	0.190	0.164	0.184	0.142	0.202	0.202	0.202	0.200

Chapter 2

A Martingale-Difference-Divergence-Based Test for Specification with Application to Gravity Models

[1]

2.1 Introduction

Since Hausman's (1978) seminal work, a large literature has been developed on testing for the correct specification of functional forms. Kernel smoothing method constitutes one of the most popular approaches towards the construction of consistent model specification tests; see, e.g., Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996), Li and Wang (1998), Horowitz and Spokoiny (2001), and Hsiao, Li, and Racine (2007) for cross-sectional data, Robinson (1989) and Fan and Li (1999) for time series data, Su and Lu (2013) and Su, Jin, and Zhang (2015) for panel data, and Su and Qu (2017) for spatial data. Sieve methods have also been adopted widely in nonparametric specification testing; see Eubank and Hart (1992), Wooldridge (1992), Hong and White (1995), de Jong (1996), Li, Hsiao, and Zinn (2003), among others. Instead of estimating the conditional mean via the kernel or sieve methods, one can also construct a consistent test based on the estimation of unconditional moment conditions which results in a class of nonsmoothing tests; see Bierens (1982, 1990), Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Delgado and Manteiga (2001), among others. Fan and Li (2000) discuss the relationship between smoothing and non-

[1] This is a co-authered work with Liangjun Su.

smoothing tests, and show that smoothing tests are more powerful than nonsmoothing tests for high frequency alternatives and less powerful for other local alternatives, while Horowitz and Spokoiny (2001) propose an adaptive rate-optimal test for regression models and suggest using several different smoothing parameters to compute a kernel-based test in order to ensure that the test has good power against both the low and high frequency alternatives.

In this paper we propose a new test for the correct specification of a parametric conditional mean model based on a variant of the *martingale difference divergence* (MDD hereafter) measure of conditional mean dependence between two random variables. In a sequence of papers, Székely, Rizzo, and Bakirov (2007), Székely and Rozzo (2009) and Székely and Rizzo (2014) propose to use distance covariance and distance correlation to measure the dependence between two random vectors which exhibit various nice properties. Such measures have been explored for feature screening in high dimensional regressions; see, e.g., Li, Zhong, and Zhu (2012). When one of the two random variables is scalar, Shao and Zhang (2014, SZ hereafter) propose to use MDD to measure the conditional mean dependence of the scalar random variable given a random vector (see the definition of MDD in (2.2.4) in the next section). Like the relationship between covariance and correlation, the MDD can also be rescaled to ensure that it lies between 0 and 1, yielding the *martingale difference correlation* (MDC) measure of a scalar variable given a random vector. MDD measures the departure of the conditional mean independence between a scalar response variable and a vector of covariates, which is a natural extension of the distance correlation measure proposed by Székely, Rizzo, and Bakirov (2007). MDD and MDC have many nice properties. First, both of them are nonnegative and equal zero if and only if the scalar response variable is conditionally mean independent of the covariates. This suggests that we can propose a test for the conditional mean independence hypothesis which is widely used in econometrics and statistics. Second, both measures have a closed-form formula that is only involved with certain expectation and norm calculations so that they can be easily estimated from the data based on the sample analogue principle. Third, the measures are dimension-free in the sense that the dimension of the conditioning variable is allowed to be huge. Indeed, SZ use MDC as a method to conduct high-dimensional variable selection to screen out variables that do not contribute to the conditional mean of the response variable given the covariates.

One drawback of SZ's original MDD and MDC measure is that when they are used for variable screening, both the response variable and covariates need to be observed. Therefore, we propose a variant of MDD that is used to measure the conditional mean independence of a scalar random error term given the covariates. With this variant, we propose a new consistent test for the null hypothesis that a parametric conditional mean model is correctly specified. Under the null hypothesis, the error term from the correctly specified model is conditionally mean independent of the regressors in the model. Since the error term is not observed, we propose to estimate it from the null model and construct a test statistic based on the sample analogue of this new MDD measure. We study the asymptotic distributions of the test statistic under the null and under a sequence of Pitman local alternatives. Our test shares many nice properties that a typical nonsmoothing test might have. First, its limiting distribution under the null is a mixture of central chi-square distributions that is not asymptotically pivotal. So we propose a wild bootstrap method to obtain the bootstrap p-value or critical value. Second, our test has nontrivial asymptotic power against local alternatives converging to the null at the usual parametric rate. More importantly, our test is free of the choice of any smoothing parameter (e.g., the bandwidth in kernel-based tests or the number of sieve approximating terms in sieve-based tests) and it does not suffer from the curse of dimensionality associated with kernel- or sieve-based tests. In principle, our test works for any finite dimensional regression problem where the number of covariates, q, can be huge. But for the derivation of our asymptotic distribution theory, we still need restrict q to be fixed. We conduct some Monte Carlo simulations and compare our test with some popular tests in the literature. Our simulation results indicate that our MDD-based test generally outperforms its competitors, especially for the case of high-frequency alternatives and for the case of many covariates. To the best of our knowledge, this paper is the first to consider consistent model specification test in the presence of many covariates.

As an illustration, we apply our test to test for the correct specification of functional forms in gravity equations that are frequently used to model the bilateral trade flow between two countries/regions. Most of the empirical studies use the log-linearized model that implies constant elasticity of trade. In an influential paper Santos Silva and Tenreyro (2006) raise several problems associated with the log gravity equation. In particular, they study how the bias arises in the OLS estimation of the log model and find strong evidence that estimation methods based on the log-linearization of the gravity equation suffer from severe misspecification. They argue that the gravity equations should be estimated in their multiplicative form and propose the Poisson pseudo-maximum-likelihood (PPML) estimator based on the level model. We apply out test to test the functional form in both the original level equation and the log-linearized model by using four datasets. For all the datasets, we reject the log and level model coherently at 10% significance level. However, its competitors show mixed testing results for different datasets. The findings reveal the advantages of our test.

The rest of the paper is organized as follows. We introduce the hypotheses and the test statistic in Section 2. We study the asymptotic distributions of the test statistic under the null hypothesis and under a sequence of Pitman local alternatives in Section 3. We compare the MDD test with several popular tests through Monte Carlo simulations in Section 4. Section 5 provides an empirical example to illustrate the choice of functional forms in gravity equation models. Section 6 concludes. The proofs of all results are relegated to the Appendix.

Throughout the paper, we adopt the following notation. For any matrix or vector A, $\|A\|$ denotes its Euclidean norm. The operator $\xrightarrow{p}$ denotes convergence in probability and $\xrightarrow{d}$ denotes convergence in distribution.

2.2 The Hypotheses and Statistic

In this section we state the hypotheses and introduce the test statistic.

2.2.1 The Hypotheses

We consider the following parametric regression model

$$Y_i = g(X_i; \beta) + \varepsilon_i,\ \ i = 1, \ldots, n, \tag{2.2.1}$$

where Y_i is a scalar dependent variable, X_i is a $q \times 1$ vector of covariates, β is a $d \times 1$ vector of unknown parameters, and ε_i is the unobserved error term. We assume that the functional form of $g(\cdot;\cdot)$ is known up to the finite dimensional parameter β. We are interested in testing the correct specification of $g(\cdot;\cdot)$. That is, we test the null hypothesis

$$\mathbb{H}_0 :\ \ P\left\{E(Y_i|X_i) = g(X_i;\beta)\right\} = 1 \text{ for some } \beta \in \mathcal{B} \tag{2.2.2}$$

versus the alternative hypothesis

$$\mathbb{H}_1 :\ \ P\left\{E(Y_i|X_i) = g(X_i;\beta)\right\} < 1 \text{ for all } \beta \in \mathcal{B}, \tag{2.2.3}$$

where $\mathcal{B}$ is the parameter space.

2.2.2 Test Statistic

To motivate our test statistic, we follow SZ and consider the MDD of ε given X whose square is defined by

$$\text{MDD}\left(\varepsilon|X\right)^2 = \int_{\mathbb{R}^q} \left|\mathbb{E}\left[\varepsilon \exp(\mathbf{i}s'X)\right] - \mathbb{E}\left(\varepsilon\right)\mathbb{E}\left[\exp(\mathbf{i}s'X)\right]\right|^2 W(s)ds, \tag{2.2.4}$$

where $\mathbf{i} = \sqrt{-1}$, $W(s) = \frac{1}{c_q ||s||^{(1+q)}}$, $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$, and $\Gamma(\cdot)$ is the complete gamma function: $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)\, dt$. Let $(\varepsilon^\dagger, X^\dagger)$ be an independent copy of (ε, X). By Theorem 1 in SZ, we have

$$\text{MDD}(\varepsilon|X)^2 = -\mathbb{E}\left\{[\varepsilon - \mathbb{E}(\varepsilon)]\left[\varepsilon^\dagger - \mathbb{E}(\varepsilon^\dagger)\right]\left\|X - X^\dagger\right\|\right\}, \tag{2.2.5}$$

and $\text{MDD}(\varepsilon|X)^2 = 0$ if and only if $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon)$.

In our setup, ε denotes the error term in a regression such that $\mathbb{E}(\varepsilon) = 0$ is always maintained. This motivates us to consider the following variant of $\text{MDD}(\varepsilon|X)^2$

$$\text{MDD}^*(\varepsilon|X)^2 = -\mathbb{E}\left[\varepsilon\varepsilon^\dagger\left\|X - X^\dagger\right\|\right] + 2\mathbb{E}\left[\varepsilon\left\|X - X^\dagger\right\|\right]\mathbb{E}\left[\varepsilon^\dagger\right]. \tag{2.2.6}$$

The following proposition establishes the properties of $\text{MDD}^*(\varepsilon|X)^2$ that serve as the basis of our test statistic.

Proposition 2.2.1. *Let $(\varepsilon^\dagger, X^\dagger)$ be an independent copy of (ε, X), where ε is a scalar random variable and X is a $q \times 1$ random vector. Suppose that $0 < \mathbb{E}[\varepsilon^2] < \infty$ and $0 < \mathbb{E}[\|X\|^2] < \infty$. Then*

(i) MDD$^*(\varepsilon|X)^2 \geq 0$;

(ii) MDD$^*(\varepsilon|X)^2 = 0$ if and only if $\mathbb{E}(\varepsilon|X) = 0$ almost surely (a.s.).

An important implication of Proposition 2.2.1 is that we can test (2.2.2) by testing whether $\text{MDD}^*(\varepsilon_i|X_i)^2 = 0$, where $\varepsilon_i = Y_i - g(X_i;\beta_0)$. In practice, ε_i is not observed. We propose to estimate the model (2.2.1) by the nonlinear least squares (NLS) to obtain the NLS estimator $\hat{\beta}$ of β. Let $\hat{\varepsilon}_i = Y_i - g(X_i;\hat{\beta})$. We propose to estimate $n\text{MDD}^*(\varepsilon|X)^2$ by the following object

$$T_n = -\frac{1}{n}\sum\sum_{1\leq i\neq j\leq n}\hat{\varepsilon}_i\hat{\varepsilon}_j\kappa_{i,j} + \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}\hat{\varepsilon}_i\kappa_{i,j}\frac{1}{n}\sum_{k=1}^n\hat{\varepsilon}_k \tag{2.2.7}$$

where $\kappa_{i,j} \equiv \|X_i - X_j\|$. In the special case where $g(X_i;\beta)$ is linear in X_i and β, i.e., $g(X_i;\beta) = (1, X_i')\beta$, we have $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ and

$$T_n = -\frac{1}{n}\sum\sum_{1\leq i\neq j\leq n}\hat{\varepsilon}_i\hat{\varepsilon}_j\kappa_{i,j} \equiv T_n^\ell. \tag{2.2.8}$$

Other than this case, $\sum_{i=1}^n \hat{\varepsilon}_i$ is generally nonzero and second term in (2.2.7) is necessary.

Remark 1. Interestingly, $\text{MDD}(\varepsilon|X)^2$ in (2.2.4) is closely related to Bierens' (1982) and Bierens and Ploberger's (1997) integrated conditional moment (ICM) test that takes the form

$$B = \int_{\mathbb{R}^q}\left|\mathbb{E}\left[\varepsilon\exp(\mathbf{i}s'\Phi(X))\right]\right|^2 W_B(s)ds, \tag{2.2.9}$$

where $W_B(\cdot)$ is a nonnegative weight function and $\Phi(\cdot):\mathbb{R}^q\to\mathbb{R}^q$ is a smooth function. But this test requires the delicate choices of both W_B and Φ and may not be tractable in practice; see Bierens (1990) and Bierens and Ploberger (1997). When $\mathbb{E}(\varepsilon)=0$, we can also write B as

$$B^* = \int_{\mathbb{R}^q} \left|\mathbb{E}\left[\varepsilon \exp(\mathbf{i}s'\Phi(X))\right] - \mathbb{E}(\varepsilon)\,\mathbb{E}\left[\exp(\mathbf{i}s'\Phi(X))\right]\right|^2 W_B(s)ds. \tag{2.2.10}$$

Apparently, $B^*=$MDD$(\varepsilon|X)^2$ by choosing $\Phi(X)=X$ and $W_B(s)=W(s)$. In this case, we can regard MDD$(\varepsilon|X)^2$ as a special example of B. As a result, our test statistic is tied closely to Bierens' ICM test.

2.3 Asymptotic Properties

In this section we study the asymptotic properties of T_n under the null hypothesis and under a sequence of Pitman local alternatives.

2.3.1 Basic Assumptions

To facilitate the study of the local power property of our test, we consider the triangular array $\{(Y_{in},X_{in},\varepsilon_{in}),\ i=1,...,n\}$. Let $Q_n(\beta)=\frac{1}{n}\sum_{i=1}^n\left[Y_{in}-g(X_{in};\beta)\right]^2$ and $Q(\beta)=\lim_{n\to\infty}\mathbb{E}\left[Y_{in}-g(X_{in};\beta)\right]^2$. Let $g_{i\beta}(\beta)\equiv\partial g(X_{in};\beta)/\partial\beta$, and $S(\beta)\equiv\lim_{n\to\infty}\mathbb{E}\left[g_{i\beta}(\beta)g_{i\beta}(\beta)'\right]$.

We make the following assumptions.

Assumption A.1. (Y_{in},X_{in}), $i=1,2,...,n$, are independently and identically distributed (IID).

Assumption A.2. The NLS estimator $\hat{\beta}$ has the following representation

$$\hat{\beta}-\beta_0 = S^{-1}\frac{1}{n}\sum_{i=1}^n g_{i\beta}\varepsilon_i + o_P\left(n^{-1/2}\right)$$

where $g_{i\beta}=g_{i\beta}(\beta_0)$ and $S=S(\beta_0)$ is positive definite. There exists a constant $C\in(0,\infty)$ such that $E\left(g_{i\beta}g_{i\beta}'\varepsilon_i^2\right)<C$.

Assumption A.3. (i) There exists a constant $C\in(0,\infty)$ such that $\mathbb{E}(\varepsilon_i^4)\le C$ and $\mathbb{E}\|X_i\|^4\le C$.

(ii) There exists a positive definite matrix H such that

$$\sup_{\beta\in N_{\varepsilon_n}(\beta_0)}\left\|\frac{1}{n}\sum_{i=1}^n\frac{\partial^2 g(X_i;\beta)}{\partial\beta\partial\beta'}-H\right\| = o_P(1)$$

where $N_\varepsilon(\beta_0) = \{\beta \in \mathbb{B} : \|\beta - \beta_0\| \le \varepsilon\}$ and $\varepsilon_n = o(1)$.

(iii) $\frac{1}{n}\sum_{i=1}^n g_{i\beta} \xrightarrow{p} S_0$, $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}(g_{i\beta}\kappa_{i,j}) \xrightarrow{p} S_1$, and $\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n g_{i\beta}g'_{j\beta}\kappa_{i,j} \xrightarrow{p} S_2$, where $S_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathbb{E}(g_{i\beta})$, $S_1 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}(g_{i\beta}\kappa_{i,j})$, and $S_2 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left(g_{i\beta}g'_{j\beta}\kappa_{i,j}\right)$.

We assume that the observations are IID in Assumption A.1 to facilitate the asymptotic analysis. We conjecture that our result below can be extended to allow for weakly dependent time series observations but restrict ourselves to IID observations for simplicity. Assumption A.2 requires $\hat{\beta}$ follow a Bahadur representation with certain well behaved influence function. One can verify A.2 under some primitive conditions given in the literature; see, e.g., Jennrich (1969), Wu (1981), and Amemiya (1985). Assumption A.3 imposes some additional conditions to study the asymptotic distribution of our test statistics. Assumption A.3(i) imposes some moment conditions for X_i and ε_i; Assumption A.3(ii) imposes uniform convergence of the gradient function in the neighborhood of β_0; Assumption A.3(iii) imposes some convergence conditions associated with $g_{i\beta}$.

Asymptotic Distribution under the Null

The following theorem reports the asymptotic distribution of T_n.

Theorem 2.3.1. *Suppose that Assumptions A.1-A.3 hold. Then under $\mathbb{H}_0$ we have*

$$T_n \xrightarrow{d} \sum_{\nu=1}^{\infty} \lambda_\nu z_\nu^2, \quad \text{as } n \to \infty$$

where z_ν's are IID $N(0,1)$, λ_ν's are the eigenvalues of the integral equation

$$\int_{-\infty}^{\infty} \varepsilon_2^2 h(X_1, X_2) f_v(X_2) dF(\xi_2) = \lambda_v f_v(X_1),$$

$\{\varepsilon_i f_\nu(X_i)\}_{\nu=1}^{\infty}$ is an orthonormal sequence of eigenfunctions, and $h(X_1, X_2)$ is defined in Equation (2.7.3) in the Appendix.

The proof of Theorem 2.3.1 is tedious and the expression for $h(X_1, X_2)$ appears complicated. Since h depends on the underlying data generating process (DGP), T_n is not asymptotically pivotal under the null and thus we cannot tabulate its critical values. In the following we will propose a bootstrap method to obtain the bootstrap p-value to make statistical inference.

Apparently, T_n shares the same type of asymptotic null distribution as the ICM test. This is not surprising given Remark 1. As mentioned, our test does not need to specify transformation function or weight function that an ICM test needs.

Local Power Analysis

To study the asymptotic local power of T_n, we consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(n^{-1/2}): \ \mathbb{E}(\varepsilon_{in}|X_{in}) = n^{-1/2}\delta(X_{in}) \ \text{ for all } i. \tag{2.3.1}$$

It is known that in general ICM tests have nontrivial power, while the nonparametric tests break down due to the slower rate of convergence than $\sqrt{n}$ of their estimators. Theorem 2.3.2 describes the asymptotic distribution of MDD test under local alternatives and shows that MDD test has nontrivial $\sqrt{n}$ local power.

Theorem 2.3.2. *Suppose Assumption A.1-A.3 hold. Then under* $\mathbb{H}_1(n^{-1/2})$*, we have*

$$T_n \xrightarrow{d} \sum_{\nu=1}^{\infty} \lambda_\nu (z_\nu + a_\nu)^2, \ \textit{as } n \to \infty$$

where $a_\nu = \lim_{n\to\infty} \mathbb{E}\left[\delta(X_{in}) f_\nu(X_{in})\right]$ *and* $f_v(\cdot)$ *is defined in Theorem 2.3.1.*

Since $\{z_\nu\}_{\nu=1}^{\infty}$ are IID $N(0,1)$, $(z_\nu + a_\nu)^2$ is stochastically larger than z_ν^2 for $a_\nu \neq 0$. This implies that our test has nontrivial asymptotic local power against local alternatives that converge to the null at rate $n^{-1/2}$. See Fan (1998) for a similar remark.

2.4 Monte Carlo Simulation

In this section we conduct a sequence of Monte Carlo simulations to evaluate the finite sample performance of our test and compare it with some existing test statistics.

2.4.1 Data Generating Processes

We consider the following data generating processes:

$$\begin{aligned}
&\text{DGP1}(m): Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji} + \sigma_i^{(m)} \varepsilon_i,\\
&\text{DGP2}(m): Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji} + n^{-1/2} \sum_{j=1}^{m} X_{ji}^2 + \sigma_i^{(m)} \varepsilon_i,\\
&\text{DGP3}(m): Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + 2\sin(mX_{1i})\sin(mX_{2i}) + \sigma_i \varepsilon_i.
\end{aligned}$$

DGP1(m) specifies m covariates and is used to evaluate the size performance of various tests. DGP2(m) specifies m covariates and is used to evaluate the local power of various tests. DGP3(m) specifies two covariates with m-dependent frequency under the alternative. We allow for conditional heteroskedasticity in all models and generate the covariates and heteroskedasticity as follows. In DGP1 and DGP2, when $m=2$, $X_1 \sim U(0,1)$, $X_2 \sim N(0,1)$, and $\sigma^{(2)} = \{0.1 + X_1 + X_2^2\}^{1/2}$; when $m=5$, $X_j \sim U(0,j)$ for $j=1,2,3$, $X_j \sim N(0,(j-3)^2)$ for $j=4,5$, and $\sigma^{(5)} = \{0.1 + \sum_{j=1}^{3} X_j + \sum_{j=4}^{5} X_j^2\}^{1/2}$; when $m=10$, $X_j \sim U(0,j)$ for $j=1,\ldots,5$, $X_j \sim N(0,(j-5)^2)$ for $j=6,\ldots,10$, and $\sigma^{(10)} = \{0.1 + \sum_{j=1}^{5} X_j + \sum_{j=6}^{10} X_j^2\}^{1/2}$; when $m=20$, $X_j \sim U(0,j)$ for $j=1\ldots,10$, $X_j \sim N(0,(j-10)^2)$ for $j=11,\ldots,20$, and $\sigma^{(20)} = \{0.1 + \sum_{j=1}^{10} X_j + \sum_{j=11}^{20} X_j^2\}^{1/2}$. In DGP3, $X_j \sim N(0,1)$ for $j=1,2$ and $\sigma = \{0.1 + X_1^2 + X_2^2\}^{1/2}$. We specify $m=1/2$, $m=1$, and $m=2$ in DGP3(m), corresponding to low-, moderate-, and high-frequency alternatives, respectively. In all cases, we generate ε_i independently from the standard normal distribution.

We will test $\mathbb{H}_0 : E(Y_i|X_i) = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ji}$ for some $(\beta_0, ..., \beta_m)$ in DGP1(m) and DGP2(m) and $\mathbb{H}_0 : E(Y_i|X_i) = \beta_0 + \sum_{j=1}^{2} \beta_j X_{ji}$ for some $(\beta_0, \beta_1, \beta_2)$ in DGP3(m).

2.4.2 Test Statistics

We will implement our test statistic T_n and denote it as MDD in the following tables. For the purpose of comparison, we consider three popular tests for the correct specification of functional form in the literature.

The first one is Zheng's (1996) and Li and Wang's (1998) residual-based test:

$$\text{Z\&LW test}: \; T_n^{Z\&LW} = \frac{1}{n(n-1)} \sum\sum_{1 \le i \ne j \le n} \frac{1}{\Pi_{l=1}^{q} h_l} K\left(\frac{X_i - X_j}{h}\right) \hat{\varepsilon}_i \hat{\varepsilon}_j,$$

where $\hat{\varepsilon}_i$ is the residual from the parametric regression under the null, q denotes the dimension of X_i, $K(\cdot)$ is a product of univariate Epanechnikov kernel, $h = (h_1, ..., h_q)'$ is a bandwidth vector, and $a/b = (a_1/b_1, ..., a_q/b_q)'$ when $a = (a_1, ..., a_q)'$ and $b = (b_1, ..., b_q)'$ are both $q \times 1$ vectors.

The second one is Härdle and Mammen's (1993, HM) test that is based on the comparison of the nonparametric estimate and the smoothed parametric estimate of the conditional mean regression function under the null:

$$\text{HM test}: \; T_n^{HM} = n \left(\Pi_{l=1}^{q} h_l\right)^{1/2} \sum_{i=1}^{n} \left[\hat{g}_h(x_i) - \mathcal{K}_{h,n} g(x_i, \hat{\beta})\right]^2,$$

where, $\mathcal{K}_{h,n}$ denotes the smoothing operator

$$\mathcal{K}_{n,h}g(x,\hat{\beta}) = \frac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) g(X_i,\hat{\beta})}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)},$$

$\hat{\beta}$ denotes the least squares estimate of the regression coefficient under the null, $\hat{g}_h(x)$ is the Nadaraya-Watson kernel estimator of $E\left(Y_i|X_i = x\right)$ by using the kernel function $K\left(\cdot\right)$ and bandwidth h.

The last one is the ICM test Bierens and Ploberger's (1997) ICM test:

$$\text{ICM test}: \ T_n^B = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\varepsilon}_j\hat{\varepsilon}_i\prod_{k=1}^{q}\exp\left\{\left[\Phi(X_{ki})+\Phi(X_{kj})\right]^2/2\right\}$$

where $\hat{\varepsilon}_i$ is the residual from the parametric regression under the null and Φ is a one-to-one mapping function from the support of X to itself: $\Phi(X_{li}) = \tan^{-1}((X_{li} - \bar{X}_l)/s_l)$, where $\bar{X}_l$ and s_l denotes the sample mean and sample standard deviation of $\{X_{li}\}_{i=1}^n$ with X_{li} being the lth component of X_i. Fan and Li (2000) also consider the above specification for the ICM test.

In all cases, we choose the bandwidth according to Silverman's rule of thumb: $h_l = 1.06s_l n^{-1/(4+q)}$ for $l = 1,...,q$. After suitable normalization, both $T_n^{Z\&LW}$ and T_n^{HM} are asymptotically standard normally distributed under the null and they can detect local alternatives converging to the null at the nonparametric rate. In contrast, the ICM test has asymptotic null distribution similar to our MDD test and it can detect local alternatives converging to the null at the usual parametric rate.

To implement all tests, we consider the wild bootstrap to obtain the bootstrap p-values despite the fact the two kernel-based tests are asymptotically $N\left(0,1\right)$ under the null. The wild bootstrap procedure is the same as that in Wu (1986) and Härdle and Mammen (1993) and the justification of its asymptotic validity is standard. See, e.g., Su, Jin, and Zhang (2015) and Su, Hoderlein, and White (2015).

We will consider various sample sizes. When we have two covariates, we let n change from 50 to 400; when we have 5 or more covariates, we let n change from 200 to 800. The number of bootstrap resamples is 400 and the number of replications is 1000 in each scenario.

2.4.3 Simulation Results

We report the simulation results in Table 1-3 for DGP1(m)-DGP3(m), respectively, where the nominal significance levels are given by 0.01, 0.05, and 0.1. Table 1 reports the empirical levels of the four tests for DGP1(m) with different numbers of covariates. The findings are interesting. First, when the number of covariates is small $(m = 2)$, all four tests perform quite well in terms of empirical level for the number of observations as small size as 50, and the empirical levels generally improve as n increases. Second, as m increases, the levels for both HM and ICM tests diminish rapidly to zero and the degeneracy of the levels does not improve when the sample size increases from 200 to 800. This indicates that either the HM test or the ICM test has severe size distortions due to the curse of dimensionality in nonparametrics. In particular, the HM test requires nonparametric estimation under the alternative. Third, both MDD and Z&LW tests perform very well unless m is too big (20) and n is small (200). As for the Z&LW test, even though it is a kernel-based nonparametric tests, it doesn't require the estimation of the regression model under the alternative. Perhaps, this explains why it is not sensitive to the number of covariates. Overall, our MDD dominates the other three tests in terms of empirical level.

Table 2 reports the empirical power for DGP2(m) when m takes different values. We summarize some findings from Table 2. First, the ICM test has reasonable power when $m = 2$. But as m increase, the ICM test does not have any power to detect local deviations from the null. It is even inferior to the two kernel-based tests (Z&LW and HM) which have power to detect local alternatives converging to the null at slower rate than $n^{-1/2}$. Second, HM test has certain power when m increases from 2 to 5 but it loses power when m increases further. This is consistent with its empirical level behavior. Third, as expected both MDD and Z&LW tests have power even in the presence of a large number of covariates. In general, our MDD test dominates the Z&LW test in terms of local empirical power. This is also consistent with the theory because our test can detect $n^{-1/2}$-local alternatives while Z&LW test can detect local alternatives converging to the null at a slower rate than $n^{-1/2}$. In sum, for the usual $n^{-1/2}$-local alternatives, our MDD test outperforms all of its competitors under investigation.

Table 3 reports the empirical power for DGP3(m) when the alternatives are at different frequencies. First, when the frequency is low $(m = 1/2)$ or moderate $(m = 1)$, all four tests have reasonable power. Second, when the frequency is low and the sample is small, the ICM test performs fairly well and it outperforms the Z&LW and HM tests. Third, the ICM test does not have power in the high-frequency case as expected. Fourth, our MDD test is almost always the best of all.

In summary, our MDD test generally has well-controlled size and it is not sensitive to the inclusion of many covariates in the regression model. It also has higher empirical power than its competitors against both local alternatives and global alternatives.

2.4.4 Testing Nonlinear Functional Form

In the previous part, we have shown the robust performance of our MDD test for the linear null hypothesis. Before implementing our test to the application of testing gravity equation, we offer some evidences for its performance in the nonlinear cases. We are attempt to obtain the simulated size and power by imitating the structure of the real data. It's worth noticing that the simulation results will only give us a glimpse of comparison between different tests rather than solid evidences. However, we can still get some useful insights.

Data Generating Process

The parent sample is from Rose (2005) which is described in Table 5 and summarized in Table 6. We try to generate covariates sample of similar structure from simple random number generators, and then generate the dependent variable by the specified model and the estimated parameters. Table 5 shows that the number of observation in Data I (Rose (2005)) is 13974. While we will only consider the cases where the simulated sample size ranges from 100 to 800. There are two reasons why we shrink the sample. The first reason is that the computation is heavy when we consider large samples; the second reason is that the dependence between covariates in the real data will reduce the information that the sample contained. Table 6 shows that in the model we have one dependent variable *trade* and fifteen independent variables. The meaning of these variables is discussed in the next section. The value of dependent variable is strictly positive. Among the fifteen covariates, we have six continuous variables and nine discrete variables. In the following simulation process, the nine discrete variables are simplified to independent Bernoulli random variables with the same mean since the incidence for *landl* and *island* equaling two are trivial (1.4% and 4.3% respectively). For the six continuous covariates, *lyi* and *lyj* are independently generated from the same distribution, and the same are *lyhi* and *lyhj*.

Figure 1 shows the histograms and the estimated kernel densities of four continuous variables: *lyi*, *lyhi*, *ldist* and *landap*. It can be seen that *lyhi* and *landap* are nearly normally distributed, *ldist* is likely to be lognormally distributed, and *lyi* is bimodal-normally distributed. Our data generating process is based on these observations. The histograms in Figure 2 show the density of parent sample which are the same as those in Figure 1. Note that the left-upper figure are rescaled. After using maximum likelihood method to fit the data, we obtain the parameters used for data generating process. *lyi (lyj)* is generated by bimodal normal density $0.85N\left(8.26, 0.97^2\right) + 0.15N\left(9.92, 0.05^2\right)$, *lyhi (lyhj)* is generated by normal density $N\left(-0.45, 2.34^2\right)$, *ldist* is generated by lognormal density $\log N\left(2.10, 0.10^2\right)$, and *landap* is generated by normal density $N(0.59, 3.65^2)$. The density functions are shown as the curves in Figure 2.

The specified model we consider is a multiplicative error model

$$Trade_i = \exp(x_i'\beta)\,\varepsilon_i,\ \ i = 1, \ldots, n$$

where ε_i is the random error independently generated from $|N(0, \sigma^2)|$, a truncated normal distribution with mean 1, where $\sigma = \sqrt{\frac{\pi e}{2}}$ and e is the Euler's number. In the following simulation, the value of n varies case by case. In Figure 3, we set $n = 400$. Figure 3 shows the density of statistics for the simulated data and the corresponding value of statistics for the parent sample. Five statistics are considered: the mean, the standard deviation, the median, the minimum, and the maximum. Instead of using the specified model, we generate the data using a slightly different model

$$Trade_i = [\exp(x_i'\beta) + n^{-1/2}\delta\exp(\tilde{x}_i'\iota)]\varepsilon_i,\ \ i = 1, \ldots, n$$

where $\tilde{x}_i$ is a subset of x_i which contains the six continuous covariates and ι is a six-dimensional column vector of ones. δ, a scalar, is used for standardizing the local alternatives. In our simulation, $\delta = \frac{std\left(\exp\left(x_i'\beta\right)\right)}{std\left(\exp\left(\tilde{x}_i'\iota\right)\right)}$. The curves represent the estimated kernel density function of corresponding statistics for $\log(Trade_i)$. The vertical straight lines are the value of corresponding statistics for the parent sample. We can see that the simulated log dependent variable is well approximated. The interesting part is that all the five critical statistics are generated with reasonable ranges in contrast to the real data. In this way, we conclude that the parent sample is well approximated by our data generating process.

Specification Testing

The null hypothesis and the local alternative hypotheses we consider are

$$\begin{aligned} \mathbb{H}_0 &: \quad \mathbb{E}(Trade_i|X_{in}) = \exp(x_i'\beta) \\ \mathbb{H}_1(n^{-1/2}) &: \quad \mathbb{E}(Trade_i|X_{in}) = \exp(x_i'\beta) + cn^{-1/2}\delta\exp(\tilde{x}_i'\iota) \end{aligned}$$

The data generating processes for the null hypotheses and the local alternative hypotheses are specified as

$$\begin{aligned} \text{DGP4} &: \quad Trade_i = \exp(x_i'\beta)\,\varepsilon_i \\ \text{DGP5(c)} &: \quad Trade_i = \left[\exp(x_i'\beta) + cn^{-1/2}\delta\exp(\tilde{x}_i'\iota)\right]\varepsilon_i \end{aligned}$$

where $\varepsilon_i \sim |N(0, \pi e/2)|$. $\tilde{x}_i$ is a subset of x_i which contains the six continuous covariates and ι is a six-dimensional column vector of ones. The number of observations $n = 100$, 200, 400, and 800. δ, a scalar, is used for standardizing the local alternatives. In our simulation, $\delta = \frac{std\left(\exp\left(x_i'\beta\right)\right)}{std\left(\exp\left(\tilde{x}_i'\iota\right)\right)}$ and $c = 1$ or 2. The simulation results are shown in Table 4. All the results are in 0.05 significance level.

From the first two columns we have that only our MDD test has the reasonable sizes for small samples, $n = 100$ and 200. When the sample is large, $n = 400$ and 800, we can see that only MDD test and Z&LW test have nontrivial power. HM test and ICM test basically fail in this case. These results coincide with the previous results of linear cases in Table 2 when we have many covariates.

2.5 Testing the Functional Form in Gravity Equations

In this section we apply our test to various datasets that are used to study the gravity equations in economics.

2.5.1 Model

Since its introduction by Tinbergen (1962), the gravity model has been widely used in international economics to explain the flows of international and subnational trade. Theoretical considerations on the proper use and deviations from the gravity model have been a topic of considerable interest in the literature; see Feenstra, Markusen, and Rose (2001), Anderson and van Woncoop (2003), Henderson and Millimet (2008), among others. Kepaptsoglou, Karlaftis, and Tsamboulas (2010) review the empirical literature on gravity models from 1999 to 2009.

Of our particular interest is the functional form specification in gravity models. Following Anderson and van Woncoop (2003) and Santos Silva and Tenreyro (2006), we use T_{ij} to denote the bilateral trade flow between country/region i and country/region j. In its simplest form, T_{ij}, is proportional to the two countries' GDPs, denoted by Y_i and Y_j, and inversely proportional to their distance, D_{ij}. More generally, we have

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3},$$

where α_0, α_1, α_2, and α_3 are unknown parameters. In practice, researchers often control other country characteristics and consider two empirical stochastic versions of T_{ij}, which are called the level model and the log model (see Henderson and Millimet (2008)):

$$\text{Level Model:} \quad \mathbb{E}(T_{ij}|Y_i, Y_j, D_{ij}, X_{ij}) = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \exp(X_{ij}'\gamma), \tag{2.5.1}$$

$$\text{Log Model:} \quad \mathbb{E}(\log T_{ij}|Y_i, Y_j, D_{ij}, X_{ij}) = \beta_0 + \beta_1 \log Y_i + \beta_2 \log Y_j + \beta_3 \log D_{ij} + X_{ij}'\gamma, \tag{2.5.2}$$

where β_0, β_1, β_2, and β_3 are unknown scalars, X_{ij} is a vector of other covariates, and γ is a unknown vector.

The linearity in the log model (2.5.2) simplifies the estimation procedure. Santos Silva and Tenreyro (2006, ST hereafter) highlight various issues associated with the gravity estimations estimated in log form. For example, the OLS estimation of the log model would be problematic when there are many zeros of T_{ij}. By omitting observations with zero values of trade, the estimates are subject to the notorious sample selection bias. In addition, the log-linear specification also generates systematic bias as a consequence of Jensen's inequality. These problems can be overcome by estimating the level equation in (2.5.1) using nonlinear estimator. ST propose the Poisson pseudo-maximum-likelihood (PPML) estimator based on the level model that becomes the new fashion in trade to estimate the gravity equation (c.f., Bosquet and Boulhol (2009)). It is shown that heteroskedasticity in the multiplicative error in the level model makes the log-model-based estimator biased. Since the model specification assumption is imperative in their analysis, they compare the PPML estimator with several other methods and apply the Ramsey's (1969) RESET method to test for the functional form. Nevertheless, it is well known that the RESET test is an inconsistent test, can only be used to test for neglected nonlinearity in linear models, and should be replaced by consistent specification tests to ensure reliable inferences.

Below we will apply our MDD test and its competitors to test the correct specification of the gravity equations in both level and log forms. Since the gravity equations can be estimated with both panel and cross-sectional models and data, one should conduct the analysis for both types of model and data. But because we have only developed our specification test theory for cross-sectional data (and the other nonparametric tests are mainly studied for cross-sectional or time series data but not for panel data), we will follow Henderson and Millimet (2008) and focus on the cross-sectional model and data below and leave the case of panel data for future research.

2.5.2 Data

We consider four datasets, which are summarized in Tables 5-9.[2]Table 5 gives brief descriptions of these datasets.[3] Tables 6-9 report the summary statistics for these four datasets, respectively.

The first dataset is used in Rose (2004) and it covers bilateral trade flow for 180 countries from 1980-2000. Martinez-Zarzoso (2013) uses the data of year 1990 to evaluate different estimation methods for the gravity models and to test the model specification with Park-type tests, which are, again, not consistent. In order to compare with the results in Martinez-Zarzoso (2013), we also use only the data for year 1990.

The second dataset is taken from ST. It covers bilateral trade flow for 136 countries in year 1990. The major difference between this dataset and the first one is that the value of bilateral trade is allowed to be zero here but not in the first dataset. Because of the zero-value of trade flow, one cannot directly log-transform the dependent variable. The general practical solution is to add a small number, like 1 in our case, to make it always positive. It is worth noticing that both the first and second datasets contain about a dozen of covariates.

The third dataset is from Glick and Rose (2002). It covers bilateral trade flow for 132 countries from 1948-1997. Henderson and Millimet (2008) use the data of year 1995 to compare the performance of the log and level models. Thus we only consider the data of year 1995 as well. It contains less number of observations and less number of covariates than the first two datasets.

The last dataset is from Millimet and Osang (2007) and it covers 96 U.S. states from 1993-1997. Henderson and Millimet (2008) also utilize this dataset. We only use the data of year 1997 that contains about two thousand observations and five covariates.

For consistency, we unify the variable names across the four datasets; *trade* denotes the level value of total exports from one country/state to the other; *lyex* (*lyim*) and *lypex* (*lypim*) denote income and income per capita in the exporter (importer) after taking log; *ldist* denotes the log of geographic distances between two districts. *border* is a dummy variable that takes value one if a common border is shared and zero otherwise; *comlang* is a dummy that is one if a common language is shared and zero otherwise; *colony* is a dummy that is one if they are colonized each other and zero otherwise; *landl* is the number of landlocked districts in the

[2]For more datasets, we refer the readers to Kepaptsoglou, Karlaftis, and Tsamboulas (2010).

[3]The first and the third datasets are both downloaded from Andrews Rose's website: *http://faculty.haas.berkeley.edu/arose/.* the second dataset is downloaded from the "Log of gavity" webpage: *http://personal.lse.ac.uk/tenreyro/LGW.html.* The fourth dataset is downloaded from the data archive of Journal of Applied Econometrics: *http://qed.econ.queensu.ca/jae/.* The authors are grateful to Reuven Glick, Daniel Henderson, Daniel Millimet, Andrews Rose, Santos Silva, and Silvana Tenreyro for making the data available.

pair; *landl_ex* (*land_im*) is a dummy that is one if the exporter (importer) is landlocked and zero otherwise; *island* is zero if they are neither islands, one if one of them is island, and two if both of them are islands; *landap* is the total area of both districts after taking log; *comfrt* is a dummy that is one when the trading partners belong the the same trade agreement and zero otherwise; *custrict* is a dummy that is one when the trading partners share a common currency and zero otherwise; *comcol* is a dummy that is one if they were ever colonies after 1945 with the same colonizer and zero otherwise; *lremo_ex* (*lremo_im*) is the log value of exporter's (importer's) remoteness as described in ST; *open* is a dummy that is one if one of them is in a preferential-trade agreement; and *home* is a dummy for intrastate trade.

2.5.3 Test Results

We implement the four model specification tests as considered in the simulation section. Table 10 show the test results for the level model and the log model.

For the log model, we summarize the findings as follows. First, both our MDD test and Z&LW test reject the log model at all conventional significance levels (0.01, 0.05, and 0.10). This is consistent with our simulation findings as both tests have well behaved size and reasonable power in various scenarios. Second, despite the low power of the HM test in the case of many covariates and the inconsistency of the RESET test, both tests also reject the log model for all four datasets. Third, the ICM test yields different conclusions for different datasets at the 5% significance level. For example, it fails to reject the log model for datasets I and IV and rejects the log model for datasets II and III. We conjecture that the failure of rejection may be due to the serious under-size distortion and low power property of this test. In sum, across the top panel of Table 10, we can conclude that the log model can be safely rejected.

For the level model, the findings are mixed. First, our MDD test reject the level model for all four datasets at 10% significance level, which shows its coherent performance among different datasets. In particular, it reject the level model for the Data I, II, and III at 1% significance level. Second, all the other three tests deliver different conclusions for different datasets. For example, the two kernel-based nonparametric tests, Z&LW test and HM test, reject the level model at the 1% level for datasets I and II and fail to reject the level model at the 5% level for datasets III and IV, but these two tests yield different conclusions for dataset IV at the 10% significance level. In addition, the ICM test fail to reject the level model at the 5% level for all four datasets but can reject the level model at the 10% level for Data III.

The findings based upon our MDD test is inconsistent with ST's findings that support the level model and the results in Henderson and Millimet (2008) and Martinez-Zarzoso (2013) that support both models, which may be due to the lack of power of their tests.

2.6 Conclusion

In this paper we have proposed a novel consistent model specification test based on the MDD of the error term given the covariates. The MDD equals zero if and only if error term is conditionally mean independent of the covariates. It does not require any nonparametric estimation under the null or alternative and is applicable even if we have many covariates in the regression model. We have established the asymptotic distributions of our test statistic under the null and under a sequence of Pitman local alternatives converging to the null at the usual parametric rate. Simulations demonstrate that our MDD test has a superb performance and generally dominates its competitors in a variety of scenarios. We apply our test to study the correct specification of functional form in gravity equations for both the level and log models. For all the datasets, we reject the log and level model coherently at 10% significance level. However, its competitors show mixed testing results for different datasets.

Several extensions are possible. First, it is easy to extend our method to test the correct specification of a semiparametric models, e.g., partially linear, additive, or single index models. In this case, one needs to estimate the semiparametric model under the null and apply undersmoothing to ensure that the bias in the semiparametric estimation is asymptotically vanishing. Second, one can extend our test to test for the correct specification of a conditional mean model in panel data models where complication arises due to the presence of unobserved individual heterogeneity. Third, we conjecture that it is also possible to extend the distance covariance or MDD to measure the dependence between two random vectors/variables conditional on a third one that is dimension-free. Recently there is a growing interest in testing conditional independence; see, e.g., Su and White (2007, 2008, 2014), Song (2009), Linton and Gozalo (2014), and Huang, Sun, and White (2016). But all of these tests are subject to the curse of dimensionality issue and are generally not applicable when the dimension of conditioning variable is large (e.g., larger than 6). So it is worthwhile to consider a dimension-free measure of conditional dependence based on which a sample analogue can be constructed and used to test for the null of conditional independence. We leave these topics for future research.

2.7 Appendix

2.7.1 Proof of the results

Proof of Proposition 2.2.1. (i) Note that

$$\begin{aligned}\text{MDD}^*(\varepsilon|X)^2 &= -\mathbb{E}\left[\varepsilon\varepsilon^\dagger\left\|X-X^\dagger\right\|\right]+2\mathbb{E}\left[\varepsilon\left\|X-X^\dagger\right\|\right]\mathbb{E}\left[\varepsilon^\dagger\right]\\ &= -\mathbb{E}\left\{[\varepsilon-\mathbb{E}(\varepsilon)]\left[\varepsilon^\dagger-\mathbb{E}\left(\varepsilon^\dagger\right)\right]\left\|X-X^\dagger\right\|\right\}+[\mathbb{E}(\varepsilon)]^2\,\mathbb{E}\left[\left\|X-X^\dagger\right\|\right]\\ &= \text{MDD}(\varepsilon|X)^2+[\mathbb{E}(\varepsilon)]^2\,\mathbb{E}\left[\left\|X-X^\dagger\right\|\right]. \qquad (2.7.1)\end{aligned}$$

The moment conditions in the proposition imply that $\text{MDD}^*(\varepsilon|X)^2$ is finite by the Cauchy-Schwarz and triangle inequalities. The first term in (2.7.1) is nonnegative by Theorem 1 in Shao and Zhang (2014) and the second term is nonnegative. It follows that $\text{MDD}^*(\varepsilon|X)^2 \geq 0$.

(ii) First, $\text{MDD}(\varepsilon|X)^2 = 0$ if and only if $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon)$ a.s. by Theorem 1 in Shao and Zhang (2014). If $\mathbb{E}(\varepsilon|X) = 0$ a.s., then $\mathbb{E}(\varepsilon) = 0$ by the law of iterated expectations and both terms in (2.7.1) are zero, implying that $\text{MDD}^*(\varepsilon|X)^2 = 0$. If $\text{MDD}^*(\varepsilon|X)^2 = 0$, we have $\text{MDD}(\varepsilon|X)^2 = [\mathbb{E}(\varepsilon)]^2 = 0$, implying that $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon) = 0$. ■

To prove Theorem 2.3.1, we suppress the dependence of $(Y_{in}, X_{in}, \varepsilon_{in})$ on n and write it simply as $(Y_i, X_i, \varepsilon_i)$. Let $\xi_i \equiv \{(X_i', \varepsilon_i)'$. Recall that $\kappa_{i,j} = \|X_i - X_j\|$, $g_{i\beta}(\beta) = \partial g(X_{in};\beta)/\partial\beta$, $g_{i\beta} = g_{i\beta}(\beta_0)$, $S(\beta) = \lim_{n\to\infty}\mathbb{E}[g_{i\beta}(\beta)g_{i\beta}(\beta)']$, $S = S(\beta_0)$, $S_0 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}(g_{i\beta})$, $S_1 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}(g_{i\beta}\kappa_{i,j})$, and $S_2 = \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}\left(g_{i\beta}g_{j\beta}'\kappa_{i,j}\right)$. Let $\mu_\kappa \equiv \mathbb{E}(\kappa_{1,2})$ and $g_{i\beta\beta}(\beta) = \frac{\partial^2 g(X_i;\beta)}{\partial\beta\partial\beta'}$. Let $\mathbb{E}_i$ denotes expectation with respect to variables indexed by i only and $\mathbb{E}_{i,j}$ denotes expectation with respect to variables indexed by i and j only. For example, $\mathbb{E}_i(\kappa_{i,j}) = \mathbb{E}(\kappa_{i,j}|X_j)$ when $i \neq j$ and $\mathbb{E}_{1,2}(\kappa_{1,i}\kappa_{2,i}) = \mathbb{E}(\kappa_{1,i}\kappa_{2,i}|X_i)$ when $i \neq 1, 2$.

Proof of Theorem 2.3.1. Since $\hat{\varepsilon}_i = Y_i - \hat{Y}_i = g(X_i;\beta) - g(X_i;\hat{\beta}) + \varepsilon_i = \varepsilon_i - r_i$ where $r_i \equiv g(X_i;\hat{\beta}) - g(X_i;\beta)$, we decompose T_n as follows

$$\begin{aligned}T_n &= -\frac{1}{n}\sum\sum_{1\leq i\neq j\leq n}(\varepsilon_i - r_i)(\varepsilon_j - r_j)\kappa_{i,j} + \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}(\varepsilon_i - r_j)\kappa_{i,j}\frac{1}{n}\sum_k(\varepsilon_k - r_k)\\ &= -\frac{1}{n}\sum\sum_{1\leq i\neq j\leq n}\varepsilon_i\varepsilon_j\kappa_{i,j} + \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}\varepsilon_i\kappa_{i,j}\frac{1}{n}\sum_k\varepsilon_k + \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}\varepsilon_i r_j\kappa_{i,j} - \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}\varepsilon_i\kappa_{i,j}\frac{1}{n}\sum_k r_k\\ &\quad -\frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}r_i\kappa_{i,j}\frac{1}{n}\sum_k\varepsilon_k - \frac{1}{n}\sum\sum_{1\leq i\neq j\leq n}r_i r_j\kappa_{i,j} + \frac{2}{n}\sum\sum_{1\leq i\neq j\leq n}r_i\kappa_{i,j}\frac{1}{n}\sum_k r_k\\ &\equiv T_{n1}+T_{n2}+T_{n3}+T_{n4}+T_{n5}+T_{n6}+T_{n7},\ \text{say.}\end{aligned}$$

Lemmas 2.7.1-2.7.7 below establish the asymptotic properties of T_{nm}, $m=1,\ldots,7$.

Lemma 2.7.1. *$T_{n1}=-n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} h_1(\xi_i,\xi_j)+o_p(1)=O_p(1)$, where $h_1(\xi_i,\xi_j)=\varepsilon_i\varepsilon_j\kappa_{i,j}$.*

Proof. $T_{n1}=-\frac{n-1}{n}\times nU_{n1}$, where $U_{n1}=\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}\psi^{(1)}(\xi_i,\xi_j)$ and $\psi^{(1)}(\xi_i,\xi_j)=\varepsilon_i\varepsilon_j\kappa_{i,j}$. Apparently, $\psi^{(1)}$ is symmetric in its two arguments, $\mathbb{E}_1\left[\psi^{(1)}(\xi_1,\xi_2)|\xi_2\right]=0$, and

$$\mathbb{E}\left[\psi^{(1)}(\xi_1,\xi_2)^2\right]=\mathbb{E}\left[\varepsilon_1^2\varepsilon_2^2\left\|X_1-X_2\right\|^2\right]\le 4\mathbb{E}\left[\varepsilon_1^2\varepsilon_2^2\left\|X_1\right\|^2\right]=4\mathbb{E}\left[\varepsilon_1^2\left\|X_1\right\|^2\right]\mathbb{E}\left[\varepsilon_2^2\right]<\infty$$

by Assumptions A.1 and A.3. So U_{n1} is a second order degenerate U-statistic satisfying the conditions of Theorem 1 in Section 3.2.2 of Lee (1990, pp. 79-80). It follows that

$$nU_{n1}\xrightarrow{d}\sum_{\nu=1}^{\infty}\lambda_\nu^{(1)}(Z_\nu^2-1),$$

where Z_ν's are IID $N(0,1)$ and $\lambda_\nu^{(1)}$'s are the eigenvalue of the integral equation $\int\psi^{(1)}(\xi_1,\xi_2)f(\xi_2)\,dF(\xi_2)$ $=\lambda f(\xi_1)$ with f and F being the probability density function (PDF) and cumulative distribution function (CDF) of ξ_i, respectively. Consequently, $T_{n1}=-n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}h_1(\xi_i,\xi_j)+$ $o_p(1)=O_p(1)$, where $h_1(\xi_i,\xi_j)=\varepsilon_i\varepsilon_j\kappa_{i,j}$. □

Lemma 2.7.2. *$T_{n2}-2\mathbb{E}\varepsilon_1^2\kappa_{1,2}=n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}h_2(\xi_i,\xi_j)+o_p(1)=O_p(1)$, where $h_2(\xi_i,\xi_j)=$ $\varepsilon_i\varepsilon_j[\mathbb{E}_1(\kappa_{i,1})+\mathbb{E}_1(\kappa_{j,1})]$.*

Proof. First, we make the following decomposition:

$$\begin{aligned}T_{n2}&=\frac{2}{n}\sum\sum_{1\le i\ne j\le n}\varepsilon_i\kappa_{i,j}\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k\\&=\frac{2}{n^2}\sum\sum_{1\le i\ne j\le n}\varepsilon_i^2\kappa_{i,j}+\frac{2}{n^2}\sum\sum_{1\le i\ne j\le n}\varepsilon_i\varepsilon_j\kappa_{ij}+\frac{2}{n^2}\sum\sum\sum_{1\le i\ne j\ne k\le n}\varepsilon_i\varepsilon_k\kappa_{i,j}\\&\equiv T_{n21}+T_{n22}+T_{n23},\ \text{say.}\end{aligned}$$

By the law of large numbers (LLN) for the second order U-statistic, $T_{n21}=2\mathbb{E}\left(\varepsilon_1^2\kappa_{1,2}\right)+$ $o_p(1)$. By Lemma 2.7.1, $T_{n22}=-\frac{2}{n}T_{n1}=O_p\left(\frac{1}{n}\right)$. Next, notice that $T_{n23}=\frac{n(n-1)(n-2)}{n^3}\times$ nU_{n2} where $U_{n2}=\binom{n}{3}^{-1}\sum\sum\sum_{1\le i<j<k\le n}\psi^{(2)}(\xi_i,\xi_j,\xi_k)$ and $\psi^{(2)}(\xi_i,\xi_j,\xi_k)=\frac{1}{3}(\varepsilon_i\varepsilon_k\kappa_{i,j}+\varepsilon_i\varepsilon_j\kappa_{i,k}+$ $\varepsilon_j\varepsilon_k\kappa_{j,k}+\varepsilon_j\varepsilon_i\kappa_{j,k}+\varepsilon_k\varepsilon_i\kappa_{j,k}+\varepsilon_k\varepsilon_j\kappa_{i,k})$. Noting that $\psi^{(2)}$ is symmetric in its three arguments, $\mathbb{E}[\psi^{(2)}(\xi_1,\xi_2,\xi_3)]=0$, $\mathbb{E}[\psi^{(2)}(\xi_1,\xi_2,\xi_3)|\xi_1]=0$, and $\mathbb{E}\left[\psi^{(2)}(\xi_1,\xi_2,\xi_3)|\xi_1,\xi_2\right]=\frac{1}{3}\varepsilon_1\varepsilon_2[(\mathbb{E}_3\left(\kappa_{1,3}\right)+$

$\mathbb{E}_3(\kappa_{2,3})] \equiv h_2^{(2)}(\xi_1,\xi_2)$. Let $h_3^{(2)}(\xi_1,\xi_2,\xi_3) = \psi^{(2)}(\xi_1,\xi_2,\xi_3) - [h_2^{(2)}(\xi_1,\xi_2) + h_2^{(2)}(\xi_1,\xi_3) + h_2^{(2)}(\xi_2,\xi_3)]$. By the Hoeffding's decomposition (e.g., Lee (1990, p.26)), we have

$$U_{n2} = 3H_{2n}^{(2)} + H_{3n}^{(2)},$$

where $H_{2n}^{(2)} = \binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_2^{(2)}(\xi_i,\xi_j)$ and $H_{3n}^{(2)} = \binom{n}{3}^{-1} \sum\sum\sum_{1\le i<j<k\le n} h_3^{(2)}(\xi_i,\xi_j,\xi_k)$. By moment calculations, $\mathbb{E}[H_{3n}^{(2)}] = 0$ and $\mathrm{Var}[H_{3n}^{(2)}] = O(n^{-3})$, implying that $H_{3n}^{(2)} = O_p(n^{-3/2})$. In addition, $H_{2n}^{(2)}$ is a standard second order degenerate U-statistic such that $nH_{2n}^{(2)} = O_p(1)$. It follows that

$$T_{n23} = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_2(\xi_i,\xi_j) + O_p(n^{-1/2}) = O_p(1),$$

where $h_2(\xi_i,\xi_j) = \varepsilon_i\varepsilon_j[\mathbb{E}_1(\kappa_{i,1}) + \mathbb{E}_1(\kappa_{j,1})]$. Combining these results, we have $T_{n2} - 2\mathbb{E}(\varepsilon_1^2\kappa_{1,2}) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_2(\xi_i,\xi_j) + o_p(1) = O_p(1)$. □

Lemma 2.7.3. *$T_{n3} - 2\mathbb{E}\left(\varepsilon_1^2\kappa_{1,2}g_{1\beta}'S^{-1}g_{2\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1\le i\neq j\le n} h_3(\xi_i,\xi_j) + o_p(1)$, where $h_3(\xi_i,\xi_j) = \varepsilon_i\varepsilon_j \times [g_{j\beta}'S^{-1}\mathbb{E}_1(g_{1\beta}\kappa_{i,1}) + g_{i\beta}'S^{-1}\mathbb{E}_1(g_{1\beta}\kappa_{j,1})]$.*

Proof. By the second order Taylor expansion,

$$r_i = g(X_i;\hat\beta) - g(X_i;\beta_0) = g_{i\beta}'(\hat\beta-\beta_0) + \frac{1}{2}(\hat\beta-\beta_0)'g_{i\beta\beta}(\breve\beta)(\hat\beta-\beta_0), \tag{2.7.2}$$

where $\breve\beta$ lies between $\hat\beta$ and β_0 elementwise. It follows that

$$\begin{aligned} T_{n3} &= \frac{2}{n}\sum\sum_{1\le i\neq j\le n}\varepsilon_i r_j\kappa_{i,j} \\ &= \left(\hat\beta-\beta_0\right)'\frac{2}{n}\sum\sum_{1\le i\neq j\le n} g_{j\beta}\varepsilon_i\kappa_{i,j} + \left(\hat\beta-\beta_0\right)'\frac{1}{n}\sum\sum_{1\le i\neq j\le n} g_{j\beta\beta}(\breve\beta)(\hat\beta-\beta_0)\varepsilon_i\kappa_{i,j} \\ &\equiv T_{n31} + T_{n32}. \end{aligned}$$

By Assumption A.1-A.3, we can readily show that $\frac{1}{n}\sum\sum_{1\le i\neq j\le n} g_{j\beta}\varepsilon_i\kappa_{i,j} = O_p\left(n^{1/2}\right)$ and $\hat\beta - \beta_0 = O_p\left(n^{-1/2}\right)$. Then we have $T_{n31} = \bar T_{n31} + o_p(1)$, where

$$\bar T_{n31} = \frac{2}{n^2}\sum_{k=1}^{n}\varepsilon_k g_{k\beta}'S^{-1}\sum\sum_{1\le i\neq j\le n} g_{j\beta}\varepsilon_i\kappa_{i,j}.$$

Next, we make the following decomposition:

$$\begin{aligned}\bar{T}_{n31} &= \frac{2}{n^2}\sum\sum_{1\le i\neq j\le n}\varepsilon_i^2 g'_{i\beta}S^{-1}g_{j\beta}\kappa_{i,j}+\frac{2}{n^2}\sum\sum_{1\le i\neq j\le n}\varepsilon_i\varepsilon_j g'_{j\beta}S^{-1}g_{j\beta}\kappa_{i,j}+\frac{2}{n^2}\sum\sum\sum_{1\le i\neq j\neq k\le n}\varepsilon_i\varepsilon_k g'_{k\beta}S^{-1}g_{j\beta}\kappa_{i,j}\\ &\equiv T_{n31,1}+T_{n31,2}+T_{n31,3}.\end{aligned}$$

Using the WLLN for U-statistics, we can readily show that $T_{n31,1} = 2\mathbb{E}\left[\varepsilon_1^2\kappa_{1,2}g'_{1\beta}S^{-1}g_{2\beta}\right] + o_p(1)$ and $T_{n32} = o_p(1)$. Next, notice that $T_{n31,3} = \frac{n(n-1)(n-2)}{n^3}\times nU_{n3}$ where $U_{n3} = \binom{n}{3}^{-1}\sum\sum\sum_{1\le i<j<k\le n}\psi^{(3)}(\xi_i,\xi$ and $\psi^{(3)}(\xi_i,\xi_j,\xi_k) = \frac{1}{3}(\varepsilon_i\varepsilon_k g'_{j\beta}S^{-1}g_{k\beta}\kappa_{i,j}+\varepsilon_i\varepsilon_j g'_{k\beta}S^{-1}g_{j\beta}\kappa_{i,k}+\varepsilon_j\varepsilon_k g'_{i\beta}S^{-1}g_{k\beta}\kappa_{i,j}+\varepsilon_j\varepsilon_i g'_{k\beta}S^{-1}g_{i\beta}\kappa_{k,j}$ $+\varepsilon_k\varepsilon_i g'_{j\beta}S^{-1}g_{i\beta}\kappa_{k,j}+\varepsilon_k\varepsilon_j g'_{i\beta}S^{-1}g_{j\beta}\kappa_{k,i})$. Using Hoeffding-decomposition method, we can readily show that $U_{n3} = \binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}h_3(\xi_i,\xi_j)+O_p(n^{-3/2})$, where $h_3(\xi_i,\xi_j) = \varepsilon_i\varepsilon_j[g_{j\beta}S^{-1}\mathbb{E}_j(g'_{j\beta}\kappa_{i,j})+ g_{i\beta}S^{-1}\mathbb{E}_i(g'_{i\beta}\kappa_{i,j})]$ for $i\neq j$. Then $T_{n31,3} = n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}h_3(\xi_i,\xi_j)+o_p(1)$ and

$$T_{n31}-2\mathbb{E}\left[\varepsilon_1^2\kappa_{1,2}g'_{1\beta}S^{-1}g_{2\beta}\right] = n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n}h_3(\xi_i,\xi_j)+o_p(1).$$

For any $\varepsilon>0$, we can apply Assumption A.3(i), the Markov inequality and the dominated convergence theorem to show that

$$\begin{aligned}P\left(\max_{1\le j\le n}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\kappa_{i,j}\right|\ge\varepsilon\right) &\le n\max_{1\le j\le n}P\left(\left|\sum_{i=1,j\neq i}^n\varepsilon_i\kappa_{i,j}\right|\ge n\varepsilon\right)\\ &\le \frac{1}{\varepsilon^4}\max_{1\le j\le n}\mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1,j\neq i}^n\varepsilon_i\kappa_{i,j}\right|^2 1\left\{\left|\sum_{i=1}^n\varepsilon_i\kappa_{i,j}\right|\ge n\varepsilon\right\}\right]\\ &= o(1).\end{aligned}$$

It follows that $\max_{1\le j\le N}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\kappa_{i,j}\right| = o_p(1)$. Then by Assumption A.3(ii) and the fact that $\hat{\beta}-\beta_0 = O_p\left(n^{-1/2}\right)$, we have

$$\begin{aligned}T_{n32} &= \left(\hat{\beta}-\beta_0\right)'\frac{1}{n}\sum\sum_{1\le i\neq j\le n}g_{j\beta\beta}(\breve{\beta})(\hat{\beta}-\beta_0)\varepsilon_i\kappa_{i,j}\\ &= \left(\hat{\beta}-\beta_0\right)'\frac{1}{n}\sum_{j=1}^n g_{j\beta\beta}(\breve{\beta})(\hat{\beta}-\beta_0)\sum_{i=1}^n\varepsilon_i\kappa_{i,j}\\ &\le \max_{1\le j\le N}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i\kappa_{i,j}\right|\left\{n\left\|\hat{\beta}-\beta_0\right\|^2\right\}\left\{\|H\|+o_P(1)\right\}\end{aligned}$$

$$= \quad o_p(1)\, O_p(1)\, O_p(1) = o_p(1)\,.$$

where the second equality follows because $\kappa_{i,j} = 0$ when $i = j$. Consequently, $T_{n3} - 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{1,2} g'_{1\beta} S^{-1} g_{2\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_3(\xi_i, \xi_j) + o_p(1)$. □

Lemma 2.7.4. *$T_{n4} + 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{1,2} g'_{1\beta} S^{-1} g_{3\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_4(\xi_i, \xi_j) + o_p(1)$, where $h_4(\xi_i, \xi_j) = -\varepsilon_i \varepsilon_j \times [g'_{j\beta} S^{-1} \mathbb{E}_{1,2}(g_{1\beta}\kappa_{i,2}) + g'_{i\beta} S^{-1} \mathbb{E}_{1,2}(g_{1\beta}\kappa_{j,2})]$.*

Proof. By (2.7.2) we have

$$\begin{aligned} T_{n4} &= -\frac{2}{n} \sum\sum_{1\le i\ne j\le n} \varepsilon_i \kappa_{i,j} \frac{1}{n} \sum_{k=1}^{n} r_k \\ &= -\frac{2}{n^2} \sum\sum_{1\le i\ne j\le n} \varepsilon_i \kappa_{i,j} \sum_{k=1}^{n} g'_{k\beta}(\hat\beta - \beta_0) - \frac{1}{n^2} \sum\sum_{1\le i\ne j\le n} \varepsilon_i \kappa_{i,j} \left(\hat\beta - \beta_0\right)' \sum_{k=1}^{n} g_{k\beta\beta}(\breve\beta)(\hat\beta - \beta_0) \\ &\equiv T_{n41} + T_{n42}. \end{aligned}$$

Noting that $\frac{1}{n^2} \sum\sum_{1\le i\ne j\le n} \varepsilon_i \kappa_{i,j} = O_p\left(n^{-1/2}\right)$, we can readily show that $T_{n41} = \bar T_{n41} + o_p(1)$, where

$$\begin{aligned} \bar T_{n41} &= -\frac{2}{n^3} \sum\sum_{1\le i\ne j\le n} \varepsilon_i \kappa_{i,j} \sum_{k=1}^{n} g'_{k\beta} S^{-1} \sum_{l=1}^{n} g_{l\beta} \varepsilon_l \\ &= -\frac{2}{n^3} \sum\sum\sum\sum_{1\le i\ne j\ne k\ne l\le n} \varepsilon_i \varepsilon_l \kappa_{i,j} g'_{k\beta} S^{-1} g_{l\beta} - \frac{2}{n^3} \sum\sum\sum_{1\le i\ne j\ne k\le n} \varepsilon_i^2 \kappa_{i,j} g'_{k\beta} S^{-1} g_{i\beta} + O_p(n^{-1/2}). \\ &\equiv T_{n41,1} + T_{n41,2} + O_p(n^{-1/2}). \end{aligned}$$

Write $T_{n41,1} = \frac{n(n-1)(n-2)(n-3)}{n^4} \times n U_{n4}$, where $U_{n4} = \binom{n}{4}^{-1} \sum\sum\sum\sum_{1\le i<j<k<l\le n} \psi^{(4)}(\xi_i, \xi_j, \xi_k, \xi_l)$, $\psi^{(4)}(\xi_i, \xi_j, \xi_k, \xi_l)$ $= -\frac{1}{12} \sum_{4!} \varepsilon_i \varepsilon_j \kappa_{i,l} g'_{k\beta} S^{-1} g_{j\beta}$, and $\sum_{4!}$ denotes the summation over all the 4! kinds of permutation of $\{i, j, k, l\}$. By Hoeffding decomposition, we can readily show that $U_{n4} = \binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_4(\xi_i, \xi_j) + O_p(n^{-3/2})$, where $h_4(\xi_i, \xi_j) = -\varepsilon_i \varepsilon_j [g'_{j\beta} S^{-1} \mathbb{E}_{1,2}(g_{1\beta}\kappa_{i,2}) + \nabla g'_i S^{-1} \mathbb{E}_{1,2}(g_{1\beta}\kappa_{j,2})]$. In addition, $T_{n41,2} = -2\mathbb{E}\left(\varepsilon_1^2 \kappa_{1,2} g'_{1\beta} S^{-1} g_{3\beta}\right) + o_p(1)$. It follows that

$$T_{n4,1} + 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{1,2} g'_{1\beta} S^{-1} g_{3\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_4(\xi_i, \xi_j) + o_p(1).$$

For T_{n42}, we have

$$T_{n42} \leq \left| \frac{1}{n^2} \sum\sum_{1 \leq i \neq j \leq n} \varepsilon_i \kappa_{ij} \right| \left\{ n \left\| \hat{\beta} - \beta_0 \right\|^2 \right\} \{ \|H\| + o_p(1) \} = O_p\left(n^{-1/2}\right) O_p(1) O_p(1) = O_p\left(n^{-1/2}\right).$$

Consequently, we have $T_{n4} + 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{1,2} g_{1\beta}' S^{-1} g_{3\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1 \leq i < j \leq n} h_4(\xi_i, \xi_j) + o_p(1)$. □

Lemma 2.7.5. $T_{n5} + 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{2,3} g_{1\beta}' S^{-1} g_{3\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1 \leq i < j \leq n} h_5(\xi_i, \xi_j) + o_p(1)$, *where* $h_5(\xi_i, \xi_j) = -\varepsilon_i \varepsilon_j (g_{i\beta} + g_{j\beta})' S^{-1} \mathbb{E}(g_{1\beta} \kappa_{1,2})$.

Proof. Note that

$$\begin{aligned} T_{n5} &= -\frac{2}{n} \sum\sum_{1 \leq i \neq j \leq n} r_i \kappa_{i,j} \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k \\ &= -\frac{2}{n} \sum\sum_{1 \leq i \neq j \leq n} g_{i\beta}'(\hat{\beta} - \beta_0) \kappa_{i,j} \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k - (\hat{\beta} - \beta_0)' \frac{1}{n} \sum\sum_{1 \leq i \neq j \leq n} g_{i\beta\beta}(\breve{\beta})(\hat{\beta} - \beta_0) \kappa_{i,j} \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k \\ &\equiv T_{n51} + T_{n52}. \end{aligned}$$

Noting that $\frac{1}{n} \sum_{k=1}^{n} \varepsilon_k = O_p\left(n^{-1/2}\right)$, we can readily show that $T_{n51} = \bar{T}_{n51} + o_p(1)$,

$$\begin{aligned} \bar{T}_{n51} &= -\frac{2}{n^3} \sum\sum_{1 \leq i \neq j \leq n} g_{i\beta}' S^{-1} \sum_{k=1}^{n} g_{k\beta} \varepsilon_k \kappa_{i,j} \sum_{l=1}^{n} \varepsilon_l \\ &= -\frac{2}{n^3} \sum\sum\sum\sum_{1 \leq i \neq j \neq k \neq l \leq n} \varepsilon_k \varepsilon_l \kappa_{i,j} g_{i\beta}' S^{-1} g_{k\beta} - \frac{2}{n^3} \sum\sum\sum_{1 \leq i \neq j \neq k \leq n} \varepsilon_k^2 \kappa_{i,j} g_{i\beta}' S^{-1} g_{k\beta} + O_p(n^{-1/2}) \\ &= T_{n51,1} + T_{n51,2} + O_p(n^{-1/2}), \end{aligned}$$

Write $T_{n51,1} = \frac{n(n-1)(n-2)(n-3)}{n^4} \times n U_{n5}$ and $U_{n5} = \binom{n}{4}^{-1} \sum\sum\sum\sum_{1 \leq i < j < k < l \leq n} \psi^{(5)}(\xi_i, \xi_j, \xi_k, \xi_l)$, $\psi^{(5)}(\xi_i, \xi_j, \xi_k, \xi_l) = -\frac{1}{12} \sum_{4!} \varepsilon_i \varepsilon_j \kappa_{k,l} \nabla g_i' S^{-1} \nabla g_k$, and $\sum_{4!}$ denotes the summation over all the 4! kinds of permutation of $\{i, j, k, l\}$. By Hoeffding decomposition, $U_{n5} = \binom{n}{2}^{-1} \sum\sum_{1 \leq i < j \leq n} h_5(\xi_i, \xi_j) + O_p(n^{-3/2})$, where $h_5(\xi_i, \xi_j) = -\varepsilon_i \varepsilon_j (g_{i\beta} + g_{j\beta})' S^{-1} \mathbb{E}(g_{1\beta} \kappa_{1,2})$. Then $T_{n51} = n\binom{n}{2}^{-1} \sum\sum_{1 \leq i < j \leq n} h_5(\xi_i, \xi_j) + o_p(1)$. In addition, $T_{n51,2} = -2\mathbb{E}\left(\varepsilon_1^2 \kappa_{2,3} g_{1\beta}' S^{-1} g_{3\beta}\right) + o_p(1)$. Therefore,

$$T_{n51} + 2\mathbb{E}\left(\varepsilon_1^2 \kappa_{2,3} g_{1\beta}' S^{-1} g_{3\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1 \leq i < j \leq n} h_5(\xi_i, \xi_j) + o_p(1).$$

Following the analysis of T_{n42}, we can readily show that $T_{n52} = O_p\left(n^{-1/2}\right)$. Consequently, the lemma follows. $\square$

Lemma 2.7.6. $T_{n6}+\mathbb{E}\left(\varepsilon_i^2 g_{i\beta}' S^{-1} S_2 S^{-1} g_{i\beta}\right) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_6(\xi_i,\xi_j)+o_p(1)$, *where* $h_6(\xi_i,\xi_j) = -\varepsilon_i\varepsilon_j \times g_{i\beta}' S^{-1} S_2 S^{-1} g_{j\beta}$.

Proof. Using (2.7.2), the fact that $\hat\beta-\beta_0 = O_p\left(n^{-1/2}\right)$ and Assumptions A.2 and A.3(iii), we can readily show that

$$
\begin{aligned}
T_{n6} = {} & -\frac{1}{n}\sum\sum_{1\le i\ne j\le n} r_i r_j \kappa_{i,j} \\
= {} & -(\hat\beta-\beta_0)'\frac{1}{n}\sum\sum_{1\le i\ne j\le n} g_{i\beta} g_{j\beta}'\kappa_{i,j}(\hat\beta-\beta_0) + O_p\left(n^{-1/2}\right) \\
= {} & -n(\hat\beta-\beta_0)' S_2 (\hat\beta-\beta_0) + o_p(1) \\
= {} & -\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g_{i\beta}' S^{-1} S_2 S^{-1} \sum_{j=1}^{n} g_{j\beta}\varepsilon_j + o_p(1) \\
= {} & -\frac{1}{n}\sum\sum_{1\le i\ne j\le n} \varepsilon_i\varepsilon_j g_{i\beta}' S^{-1} S_2 S^{-1} g_{j\beta} - \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 g_{i\beta}' S^{-1} S_2 S^{-1} g_{i\beta} + o_p(1) \\
\equiv {} & T_{n61} + T_{n62} + o_p(1),
\end{aligned}
$$

where $S_2 = \lim_{n\to\infty} \frac{1}{n^2}\sum\sum_{1\le i\ne j\le n} \mathbb{E}\left(g_{i\beta} g_{j\beta}' \kappa_{i,j}\right)$. Obviously, $T_{n61} = \frac{n(n-1)}{n^2} n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} \left(-\varepsilon_i\varepsilon_j g_{i\beta}' S^{-1} S_2 S^{-1} g\right.$ $+o_p(1) = nU_{n6}+o_p(1)$ where $U_{n6} = \binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} h_6(\xi_i,\xi_j)$ and $h_6(\xi_i,\xi_j) = -\varepsilon_i\varepsilon_j g_{i\beta}' S^{-1} S_2 S^{-1} g_{j\beta}$. Then $T_{n61} = n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} h_6(\xi_i,\xi_j) + o_p(1)$.

In addition, $T_{n62} = -\mathbb{E}\left(\varepsilon_i^2 g_{i\beta}' S^{-1} S_2 S^{-1} g_{i\beta}\right)+o_p(1)$. It follows that $T_{n6}+\mathbb{E}\left(\varepsilon_i^2 g_{i\beta}' S^{-1} S_2 S^{-1} g_{i\beta}\right)$ $= n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} h_6(\xi_i,\xi_j) + o_p(1)$. $\square$

Lemma 2.7.7. $T_{n7}-2\mathbb{E}\left(\varepsilon_i^2 g_{i\beta}' S^{-1} S_1 S_0' S^{-1} g_{i\beta}\right) = n\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} h_7(\xi_i,\xi_j)+o_p(1)$, *where* $h_7(\xi_i,\xi_j) = \varepsilon_i\varepsilon_j \left(g_{i\beta}' S^{-1} S_1 S_0' S^{-1} g_{j\beta} + g_{j\beta}' S^{-1} S_1 S_0' S^{-1} g_{i\beta}\right)$.

Proof. Using (2.7.2) and the fact that $\hat\beta-\beta_0 = O_p\left(n^{-1/2}\right)$, we can readily show that

$$
\begin{aligned}
T_{n7} = {} & \frac{2}{n}\sum\sum_{1\le i\ne j\le n} r_i \kappa_{i,j} \frac{1}{n}\sum_{k=1}^{n} r_k \\
= {} & 2(\hat\beta-\beta_0)'\left\{\frac{1}{n^2}\sum\sum_{1\le i\ne j\le n} g_{i\beta}\kappa_{i,j}\right\}\sum_{k=1}^{n} g_{k\beta}'(\hat\beta-\beta_0) + O_p\left(n^{-1/2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 2(\hat{\beta}-\beta_0)' S_1 \sum_{k=1}^{n} g'_{k\beta}(\hat{\beta}-\beta_0) + o_p(1) \\
&= \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i g'_{i\beta} S^{-1} S_1 \left(\frac{1}{n}\sum_{k=1}^{n} g'_{k\beta}\right) S^{-1} \sum_{j=1}^{n} g_{j\beta}\varepsilon_j + o_p(1) \\
&= \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i g'_{i} S^{-1} S_1 S'_0 S^{-1} \sum_{j=1}^{n} g_{j\beta}\varepsilon_j + o_p(1) \\
&= \frac{2}{n}\sum_{i=1}^{n} g'_{i\beta}\varepsilon_i S^{-1} S_1 S'_0 S^{-1} \sum_{j=1}^{n} g_{j\beta}\varepsilon_j + o_p(1) \\
&= \frac{2}{n}\sum\sum_{1\le i\ne j\le n} g'_{i\beta}\varepsilon_i S^{-1} S_1 S'_0 S^{-1} g_{j\beta}\varepsilon_j + \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i^2 g'_{i\beta} S^{-1} S_1 S'_0 S^{-1} g_{i\beta} + o_p(1) \\
&\equiv T_{n71} + T_{n72} + o_p(1).
\end{aligned}
$$

where $S_1 = \lim_{n\to\infty} \frac{1}{n^2} \sum\sum_{1\le i\ne j\le n} \mathbb{E}\left(g_{i\beta}\kappa_{i,j}\right)$ and $S_0 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(g_{i\beta}\right)$. Apparently, $T_{n71} = \frac{n(n-1)}{n^2} \times nU_{n7} + o_p(1) = nU_{n7} + o_p(1)$ and $T_{n72} = 2\mathbb{E}\left(\varepsilon_i^2 g'_{i\beta} S^{-1} S_1 S'_0 S^{-1} g_{i\beta}\right) + o_p(1)$, where $U_{n7} = \binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_7(\xi_i,\xi_j)$, and $h_7(\xi_i,\xi_j) = \varepsilon_i\varepsilon_j(g'_{i\beta} S^{-1} S_1 S'_0 S^{-1} g_{j\beta} + g'_{j\beta} S^{-1} S_1 S'_0 S^{-1} g_{i\beta})$. It follows that $T_{n7} - 2\mathbb{E}(\varepsilon_i^2 g'_{i\beta} S^{-1} S_1 S'_0 \times S^{-1} g_{i\beta}) = n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h_7(\xi_i,\xi_j) + o_p(1)$. □

Combining the results in Lemmas 2.7.1-2.7.7, we have

$$
T_n = B_1 + n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} \sum_{m=1}^{7} h_m(\xi_i,\xi_j),
$$

where each h_m is the kernel function of a second-order degenerate U-statistic. Thus $n\binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} \sum_{m=1}^{7}$ is a second-order degenerate U-statistic with kernel $\sum_{m=1}^{7} h_m(\xi_i,\xi_j)$. Then by Dunford and Schwartz (1963. p. 1087) and Theorems 2.1 and 2.3 in Gregory (1977), we have

$$
T_n - B_1 \xrightarrow{d} \sum_{\nu=1}^{\infty} \tilde{\lambda}_\nu (z_\nu^2 - 1) \text{ and } T_n \xrightarrow{d} \sum_{\nu=1}^{\infty} \tilde{\lambda}_\nu z_\nu^2,
$$

where $B_1 = \sum_{m=1}^{7} \mathbb{E}\left[h_m(\xi_i,\xi_i)\right]$, z_ν's are IID $N(0,1)$, and $\tilde{\lambda}_\nu$'s are the eigenvalues of the integral function for the following eigenvalue problem:

$$
\int_{-\infty}^{\infty} \sum_{m=1}^{7} h_m(\xi_1,\xi_2)\tilde{f}(\xi_2)dF(\xi_2) = \lambda\, \tilde{f}(\xi_1).
$$

Here, $\sum_{m=1}^{7} h_m(\xi_1, \xi_2) = \varepsilon_1\varepsilon_2 h(X_1, X_2)$ where

$$\begin{aligned} h(X_1, X_2) &= -\kappa_{1,2} + [\mathbb{E}_3(\kappa_{1,3}) + \mathbb{E}_3(\kappa_{2,3})] + g'_{2\beta}S^{-1}\mathbb{E}_3(g_{3\beta}\kappa_{1,3}) + g'_{1\beta}S^{-1}\mathbb{E}_3(g_{3\beta}\kappa_{2,3})] \\ &\quad -g'_{2\beta}S^{-1}\mathbb{E}_{3,4}(g_{3\beta}\kappa_{1,4}) - g'_{1\beta}S^{-1}\mathbb{E}_{3,4}(g_{3\beta}\kappa_{2,4}) - (g_{1\beta} + g_{2\beta})'S^{-1}S_2\mathbb{E}\left(g_{3\beta}\kappa_{3,4}\right) \\ &\quad -g'_{1\beta}S^{-1}S_2S^{-1}g_{2\beta} + g'_{1\beta}S^{-1}S_2S'_0S^{-1}g_{2\beta} + g'_{2\beta}S^{-1}S_2S'_0S^{-1}g_{1\beta}. \end{aligned} \tag{2.7.3}$$

We have

$$\begin{aligned} \lambda\tilde{f}(\xi_1) = & \int_{-\infty}^{\infty}\sum_{m=1}^{7} h_m(\xi_1, \xi_2)\tilde{f}(\xi_2)dF(\xi_2) \\ = & \int_{-\infty}^{\infty} \varepsilon_1\varepsilon_2 h(X_1, X_2)\tilde{f}(\xi_2)dF(\xi_2) \\ = & \varepsilon_1\int_{-\infty}^{\infty} \varepsilon_2 h(X_1, X_2)\tilde{f}(\xi_2)dF(\xi_2) \end{aligned}$$

Then we can write $\tilde{f}(\xi_1) = \varepsilon_1 f(X_1)$ by properly choosing $f(\cdot)$. Similarly, $\tilde{f}(\xi_2) = \varepsilon_2 f(X_2)$.

The integration equation can be rewritten as

$$\int_{-\infty}^{\infty} \varepsilon_2^2 h(X_1, X_2)f(X_2)dF(\xi_2) = \lambda f(X_1)$$

This completes the proof of Theorem 2.3.1. ■

Proof of Theorem 2.3.2. Let $\alpha_n = n^{-1/2}$. Note that

$$\begin{aligned} \hat{\varepsilon}_i = & \varepsilon_i + g(X_i;\beta) - g(X_i;\hat{\beta}) \\ = & \varepsilon_i + g'_{i\beta}(\hat{\beta} - \beta_0) + \frac{1}{2}(\hat{\beta} - \beta_0)'g_{i\beta\beta}(\breve{\beta})(\hat{\beta} - \beta_0), \end{aligned}$$

where $\mathbb{E}(\varepsilon_i|X_i) = \alpha_n\delta(X_i)$ under $\mathbb{H}_1(a_n)$. Note that the NLS estimator $\hat{\beta}$ is also $\sqrt{n}$-consistent under $\mathbb{H}_1(n^{-1/2})$, we can readily follow the proof of Theorem 2.3.1 and show that

$$\tilde{T}_n - B_1 = \frac{1}{n}\sum\sum_{1\le i\ne j\le n}\varepsilon_i\varepsilon_j h(X_i, X_j) + o_p(1)$$

under $\mathbb{H}_1(n^{-1/2})$. The conclusion then follows from Theorem 2.3 in Gregory (1977). ■

Table I. Empirical Size under the Null Hypothesis

	Level	0.1				0.05				0.01			
DGP	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	IMD	Z&LW	HM	ICM
DGP1(2)	50	0.130	0.136	0.126	0.123	0.050	0.062	0.060	0.045	0.006	0.010	0.007	0.006
	100	0.119	0.117	0.115	0.116	0.060	0.063	0.054	0.054	0.012	0.013	0.011	0.006
	200	0.103	0.102	0.125	0.105	0.052	0.052	0.059	0.050	0.015	0.009	0.008	0.009
	400	0.096	0.091	0.096	0.102	0.039	0.040	0.041	0.052	0.009	0.008	0.008	0.007
DGP1(5)	200	0.131	0.125	0.079	0.021	0.067	0.054	0.040	0.006	0.021	0.008	0.005	0.000
	400	0.109	0.107	0.091	0.033	0.053	0.059	0.035	0.005	0.007	0.015	0.005	0.000
	800	0.108	0.099	0.103	0.038	0.057	0.047	0.045	0.008	0.009	0.008	0.006	0.000
DGP1(10)	200	0.113	0.129	0.003	0.000	0.047	0.071	0.000	0.000	0.013	0.017	0.000	0.000
	400	0.113	0.105	0.000	0.000	0.056	0.052	0.000	0.000	0.010	0.009	0.000	0.000
	800	0.092	0.096	0.000	0.000	0.040	0.049	0.000	0.000	0.009	0.009	0.000	0.000
DGP1(20)	200	0.202	0.183	0.006	0.000	0.083	0.098	0.000	0.000	0.007	0.023	0.000	0.000
	400	0.113	0.142	0.000	0.000	0.051	0.061	0.000	0.000	0.005	0.013	0.000	0.000
	800	0.120	0.125	0.000	0.000	0.052	0.062	0.000	0.000	0.007	0.010	0.000	0.000

Table II. Empirical Power under the Local Alternatives

	Level	0.1				0.05				0.01			
DGP	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM
DGP2(2)	50	0.759	0.679	0.674	0.529	0.651	0.554	0.577	0.289	0.285	0.229	0.253	0.063
	100	0.812	0.726	0.749	0.570	0.724	0.636	0.640	0.395	0.395	0.308	0.354	0.090
	200	0.860	0.751	0.790	0.640	0.718	0.612	0.669	0.418	0.491	0.392	0.444	0.145
	400	0.846	0.772	0.799	0.641	0.746	0.660	0.691	0.445	0.501	0.396	0.449	0.185
DGP1(5)	200	0.801	0.634	0.448	0.109	0.692	0.501	0.295	0.025	0.454	0.259	0.115	0.001
	400	0.837	0.640	0.514	0.132	0.752	0.512	0.367	0.049	0.525	0.295	0.147	0.001
	800	0.838	0.612	0.474	0.135	0.743	0.490	0.343	0.051	0.516	0.262	0.160	0.006
DGP1(10)	200	0.979	0.880	0.038	0.000	0.953	0.786	0.006	0.000	0.820	0.522	0.000	0.000
	400	0.987	0.886	0.025	0.000	0.969	0.811	0.003	0.000	0.890	0.580	0.000	0.000
	800	0.977	0.834	0.023	0.000	0.961	0.757	0.001	0.000	0.873	0.548	0.000	0.000
DGP1(20)	200	0.709	0.420	0.006	0.000	0.545	0.283	0.000	0.000	0.248	0.092	0.000	0.000
	400	0.643	0.345	0.000	0.000	0.523	0.218	0.000	0.000	0.283	0.088	0.000	0.000
	800	0.611	0.325	0.000	0.000	0.492	0.216	0.000	0.000	0.279	0.075	0.000	0.000

Table III. Empirical Power under Alternatives with Different Frequencies

	Level	0.1				0.05				0.01			
DGP	n	MDD	Z&LW	HM	ICM	MDD	Z&LW	HM	ICM	IMD	Z&LW	HM	ICM
DGP3(1/2)	50	0.758	0.583	0.504	0.715	0.613	0.433	0.367	0.550	0.286	0.156	0.155	0.237
	100	0.962	0.865	0.806	0.906	0.901	0.764	0.687	0.833	0.688	0.493	0.421	0.548
	200	0.997	0.986	0.978	0.983	0.997	0.971	0.955	0.969	0.970	0.907	0.840	0.893
	400	1.000	1.000	0.999	0.999	1.000	0.999	1.000	0.998	0.999	0.998	0.994	0.993
DGP3(1)	50	0.984	0.792	0.900	0.848	0.964	0.675	0.820	0.768	0.852	0.439	0.615	0.603
	100	0.999	0.932	0.984	0.913	0.995	0.872	0.970	0.837	0.990	0.794	0.935	0.783
	200	1.000	0.988	1.000	0.973	1.000	0.978	1.000	0.949	1.000	0.955	0.997	0.893
	400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.998	1.000	0.988
DGP3(2)	50	0.766	0.184	0.489	0.170	0.600	0.117	0.307	0.091	0.226	0.041	0.109	0.013
	100	0.951	0.209	0.743	0.154	0.904	0.136	0.603	0.064	0.654	0.074	0.301	0.012
	200	1.000	0.318	0.974	0.142	0.997	0.213	0.933	0.069	0.982	0.120	0.732	0.010
	400	1.000	0.424	1.000	0.169	1.000	0.343	1.000	0.094	1.000	0.278	0.991	0.015

Table IV. Empirical Size and Power of Nonlinear Cases in 5% significance level

DGP	n	MDD	Z&LW	HM	ICM
DGP4	100	0.072	0.084	0.098	0.082
	200	0.640	0.268	0.116	0.080
	400	0.626	0.224	0.064	0.030
	800	0.618	0.214	0.066	0.026
DGP5(1)	100	0.130	0.144	0.190	0.168
	200	0.650	0.306	0.234	0.194
	400	0.604	0.180	0.058	0.024
	800	0.610	0.188	0.048	0.022

Table V. Datasets Description

No.	Reference	Obs	Object	Year
I	Rose (2005)	13974	180 Countries	1990
II	Santos Silva and Tenreyro (2006)	18360	136 Countries	1990
III	Glick and Rose (2002)	4315	132 Countries	1995
IV	Millimet and Osang (2007)	2091	96 U.S. States	1997

Table VI. Summary Satistics for Data I

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
trade	13974	2.06E+08	1.93E+09	2.40E+06	0.0000134	8.68E+10
lyi	13974	8.512	1.075	8.585	6.202	10.183
lyj	13974	8.512	1.075	8.585	6.202	10.183
lyhi	13974	-0.454	2.341	-0.792	-6.454	5.424
lyhj	13974	-0.454	2.341	-0.792	-6.454	5.424
ldist	13974	8.232	0.797	8.434	3.684	9.422
border	13974	0.024	0.154	0	0	1
comlang	13974	0.226	0.419	0	0	1
colony	13974	0.017	0.130	0	0	1
landl	13974	0.245	0.462	0	0	2
island	13974	0.398	0.571	0	0	2
landap	13974	0.593	3.653	1.213	-15.907	9.170
regional	13974	0.018	0.134	0	0	1
custrict	13974	0.010	0.098	0	0	1
comcol	13974	0.116	0.320	0	0	1

Table VII. Summary Statistics for Data II

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
trade	18360	1.72E+05	1.83E+06	17	0	1.01E+08
lypex	18360	15.744	1.893	15.891	10.646	20.854
lypim	18360	15.744	1.893	15.891	10.646	20.854
lyex	18360	7.505	1.640	7.311	4.608	10.735
lyim	18360	7.505	1.640	7.311	4.608	10.735
ldist	18360	8.786	0.742	8.954	4.877	9.899
border	18360	0.020	0.139	0	0	1
comlang	18360	0.210	0.407	0	0	1
colony	18360	0.170	0.376	0	0	1
landl_ex	18360	0.154	0.361	0	0	1
landl_im	18360	0.154	0.361	0	0	1
lremo_ex	18360	8.947	0.264	8.947	8.491	9.604
lremo_im	18360	8.947	0.264	8.947	8.491	9.604
comfrt	18360	0.025	0.156	0	0	1
open	18360	0.564	0.496	1	0	1

Table VIII. Summary Statistics for Data III

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
trade	7640	1.84E+06	1.75E+07	1.93E+04	6.33E-04	8.79E+08
lremo_ex	7640	24.771	2.144	25.090	18.231	29.238
lremo_im	7640	23.378	2.034	23.276	18.231	28.530
ldist	7640	8.167	0.799	8.354	3.783	9.422
custrict	7640	0.011	0.102	0	0	1
comlang	7640	0.199	0.400	0	0	1
comfrt	7640	0.021	0.143	0	0	1
border	7640	0.025	0.157	0	0	1
landl	7640	0.314	0.513	0	0	2
island	7640	0.355	0.546	0	0	2

Table IX. Summary Statistics for Data IV

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
trade	2091	3.19E+03	1.57E+04	6.57E+02	0.984	4.81E+05
ldist	2091	6.807	0.798	6.943	2.944	8.076
lremo_ex	2091	-0.070	0.189	-0.071	-0.469	0.417
lremo_im	2091	-0.068	0.191	-0.070	-0.469	0.417
border	2091	0.125	0.331	0	0	1
home	2091	0.023	0.150	0	0	1

Table X. Test Results: P-values

Model	Test	Data I	Data II	Data III	Data IV
Log	MDD	0.0000	0.0000	0.0000	0.0000
	Z&LW	0.0000	0.0000	0.0000	0.0000
	HM	0.0000	0.0000	0.0000	0.0000
	ICM	0.0725	0.0100	0.0175	0.0500
	RESET	0.0000	0.0000	0.0000	0.0000
Level	MDD	0.0000	0.0000	0.3700	0.0725
	Z&LW	0.0000	0.0013	0.6650	0.1850
	HM	0.0000	0.0100	0.6000	0.0775
	ICM	0.1125	0.7875	0.0625	0.1300

2.7.2 Tables and Figures

Figure 2.1. Parent Sample of Covariates

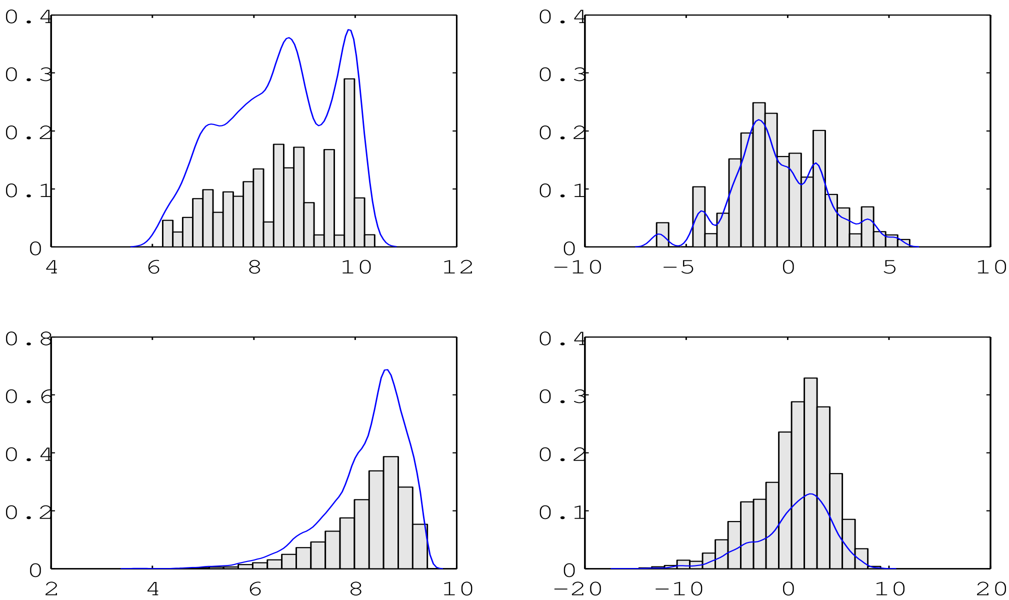


Figure 2.2. Simulated Data of Covariates

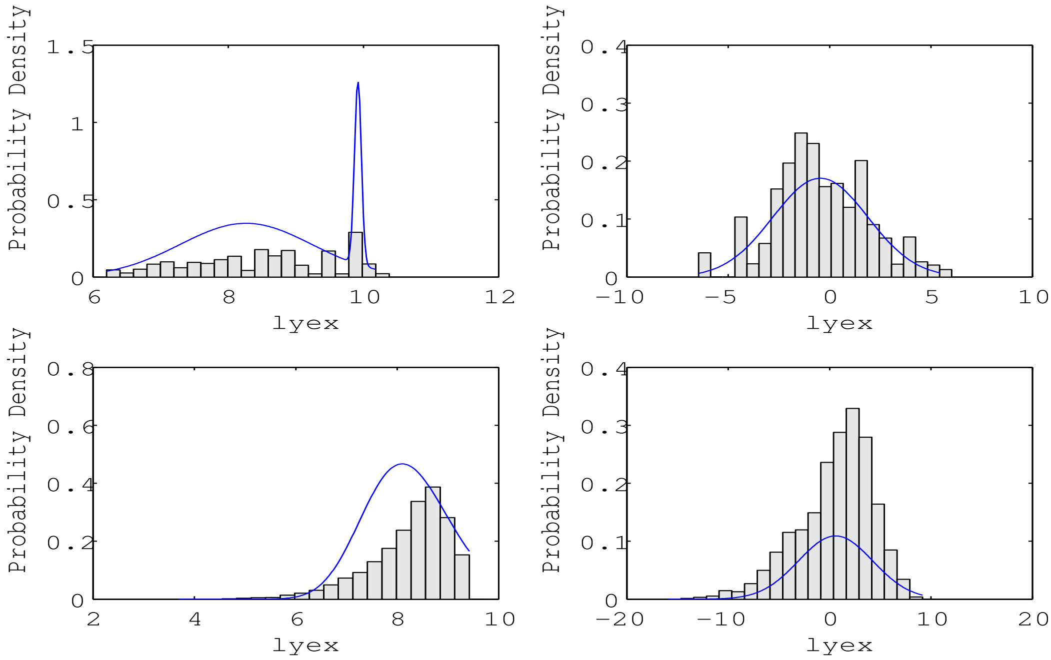
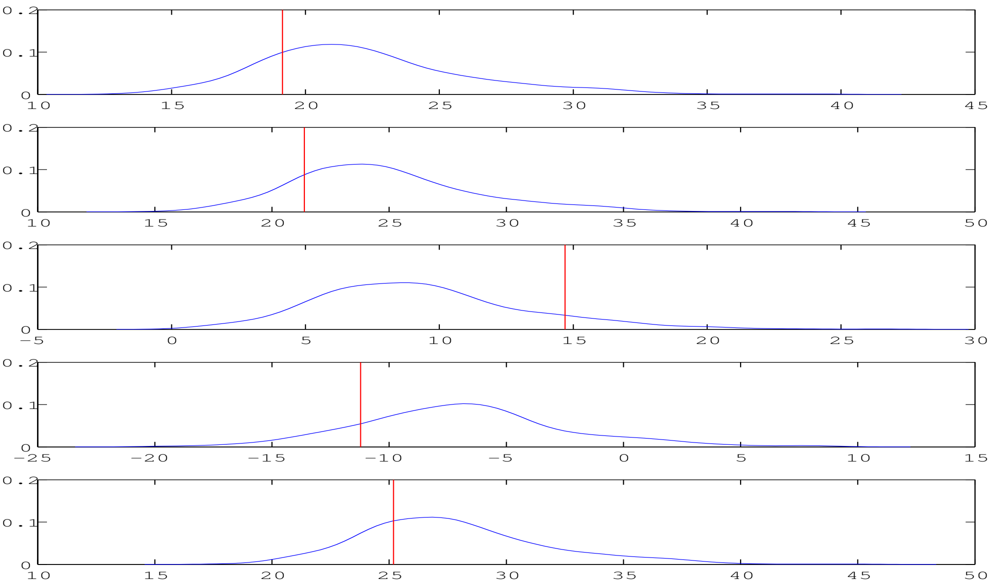


Figure 2.3. Comparison: Parent Sample and Simulated Data



Chapter 3

Estimation of Dynamic Multilevel Panel Data Models with Fixed Effects

[1]

3.1 Introduction

With the rapid revolution of big data, traditional two-dimensional panel data models sometimes cannot fully describe the characteristics of sample structures. Consider the case of observation station of air pollution (see Antweiler (2001)). The data of air pollution has three-dimensions: time, country and city, in which the latter two are obtained from the location of observation stations. Many panel data samples could be grouped in this way. Apart from the nested models as the example of air pollution, nonnested multilevel data appear everywhere as well. Bilateral trade can be modelled by a panel data model with three dimensions, which are time, importer indicator and exporter indicator (see Matyas (1997)). An excellent work considering the estimation of multilevel fixed effects panel data models is Balazsi et al (2018) (BMW hereafter). They proposed estimators for six panel data models with different kinds of fixed effects and compared them to each other. They also provided the Nickell Biases for those typical dynamic models. Nickell biases are defined as the asymptotic biases in the dynamic panel data models (see Nickell (1981)). The inconsistency is produced by the so-called incidental-parameter problem. In order to obtain consistent estimators, BMW constructed the Arellano-Bonds GMM estimator (see Arellano and Bond (1991)) for a multilevel AR(1) panel data model. In a different way, Dhaene and Jochmans (2015) proposed a split-sample jackknife estimation method to handle the Nickell biases problem in dynmic

[1] This is a co-authered work with Liangjun Su.

panel data models. One of the contributions of this paper is that we extend the jackknife estimation method to the multilevel framework.

Table 1 of BMW summarized corresponding transformations to eliminate error components for different structures. Intuitively, some models have more than one kinds of available transformations and some transformations will "over-clear" the fixed effects which leads to efficiency loss. This paper firstly (to our knowledge) proposes the expression of Nickell biases for within estimators of several popular three-dimensional dynamic panel data models and it proposes split-sample jackknife estimators, which eliminate the biases.

To clarify the ambiguity of notations, for a three-dimension variable y_{ijt}, the average along index i is denoted as $\bar{y}_{\cdot jt} = N_1^{-1}\sum_{i=1}^{N_1} y_{ijt}$. In the same way, $\bar{y}_{i\cdot t} = N_2^{-1}\sum_{j=1}^{N_2} y_{ijt}$, $\bar{y}_{ij\cdot} = T^{-1}\sum_{t=1}^{T} y_{ijt}$, $\bar{y}_{\cdot\cdot t} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} y_{ijt}$, $\bar{y}_{\cdot j\cdot} = (N_1T)^{-1}\sum_{i=1}^{N_1}\sum_{t=1}^{N_2} y_{ijt}$, $\bar{y}_{i\cdot\cdot} = (N_2T)^{-1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} y_{ijt}$, and $\bar{y}_{\cdot\cdot\cdot} = (N_1N_2T)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} y_{ijt}$. Sometimes we need to take average over t along 1 to T for $y_{ij,t-\kappa}$, we denote $\bar{y}_{ij\cdot}^{(-\kappa)} = T^{-1}\sum_{t=1}^{T} y_{ij,t-\kappa}$, analogouly, $\bar{y}_{\cdot j\cdot}^{(-\kappa)} = (N_1T)^{-1}\sum_{i=1}^{N_1}\sum_{t=1}^{T} y_{ij,t-\kappa}$, and $\bar{y}_{\cdot\cdot\cdot}^{(-\kappa)} = (N_1N_2T)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} y_{ij,t-\kappa}$. $\mathbf{I}$ is the identity matrix, we don't specify the subscript when there doesn't exist ambiguity. $\mathbf{J}$ is the matrix with all the elements are 1 and $\bar{\mathbf{J}}_N = \mathbf{J}_N/N$. $\boldsymbol{\Psi}_0 = \begin{bmatrix} \mathbf{0} & 0 \\ \mathbf{I}_{T-1} & \mathbf{0} \end{bmatrix}$, $\boldsymbol{\Gamma}_0 = \boldsymbol{\Psi}_0\left(\mathbf{I} - \rho\boldsymbol{\Psi}_0\right)^{-1}$, $\boldsymbol{\Psi} = \mathbf{I}_{N_1N_2} \otimes \boldsymbol{\Psi}_0$ and $\boldsymbol{\Gamma} = \mathbf{I}_{N_1N_2} \otimes \boldsymbol{\Gamma}_0$. Note that our definition of $\boldsymbol{\Gamma}_0$ is different from the definition in Balazsi et al (2018).

3.2 The General Within Estimators

This section discusses and extends the results of BMW. As in BMW, they considered six dynamic panel data models with different error component structures. They are three-dimensional panel data models indexed by i, j and t, whose number of observations are N_1, N_2 and T. Note that for simplicity, we exclude unbalanced panel data models where the sample size for i_1 and i_2 could be different. Also we assume that $\mathbf{x}_{ijt}$ does not contain the terms that fixed over i, j, t, or any combination of these three dimensions. The estimation problem of β when $\mathbf{x}_{ijt}$ contains fixed terms can be solved through the similar Hausman-Taylor fashion as in the two-dimensional panel data models. (see Hausman and Taylor (1981)).

$$\begin{aligned}
&\text{Model 1 } y_{ijt} = \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt};\\
&\text{Model 2 } y_{ijt} = \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \gamma_{ij} + \lambda_t + \varepsilon_{ijt};\\
&\text{Model 3 } y_{ijt} = \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \gamma_{ij} + \varepsilon_{ijt};\\
&\text{Model 4 } y_{ijt} = \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \alpha_{jt} + \varepsilon_{ijt};
\end{aligned}$$

$$\begin{aligned}\text{Model 5 } y_{ijt} =& \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt};\\ \text{Model 6 } y_{ijt} =& \rho y_{ij,t-1} + \beta'\mathbf{x}_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} + \varepsilon_{ijt}.\end{aligned}$$

The models can be summarized as

$$\text{The General Model } \mathbf{y} = \rho\mathbf{y}_{-1} + \mathbf{X}\beta + \mathbf{D}_\nu\pi_\nu + \varepsilon,$$

where y's are specified as

$$(y_{111}, \cdots, y_{11T}, \cdots, y_{1N_21}, \cdots, y_{1N_2T}, \cdots, y_{N_111}, \cdots y_{N_11T}, \cdots, y_{N_1N_21}, \ldots, y_{N_1N_2T})',$$

whose dimension is $N_1N_2T \times 1$, $\mathbf{X}$ is of dimension $N_1N_2T \times d$, β is of dimension $d \times 1$, ε is the disturbance vector of dimension $N_1N_2T \times 1$, $\mathbf{D}_\nu$ and π_ν's are column vectors of composite fixed effects parameters.

Since the dummy matrix $\mathbf{D}_\nu$ could be of not full culumn rank in some cases. There are many ways to construct the dummy matrices with full column rank. In this paper, the expressions of dummy matrices for model 1 to model 6 are respectively (see Table 1.1 of Martyas (2017)) as follows.

$$\begin{aligned}\mathbf{D}_1 =& ((\mathbf{I}_{N_1} \otimes \iota_{N_2T}), (\iota_{N_1} \otimes \mathbf{I}_{N_2} \otimes \iota_T)^*, (\iota_{N_1N_2} \otimes \mathbf{I}_T)^*);\\ \mathbf{D}_2 =& ((\mathbf{I}_{N_1N_2} \otimes \iota_T), (\iota_{N_1N_2} \otimes \mathbf{I}_T)^*);\\ \mathbf{D}_3 =& (\mathbf{I}_{N_1N_2} \otimes \iota_T);\\ \mathbf{D}_4 =& (\mathbf{I}_{N_1} \otimes \iota_{N_2} \otimes \mathbf{I}_T);\\ \mathbf{D}_5 =& ((\mathbf{I}_{N_1} \otimes \iota_{N_2} \otimes \mathbf{I}_T), (\iota_{N_1} \otimes \mathbf{I}_{N_2T})^*);\\ \mathbf{D}_6 =& ((\mathbf{I}_{N_1N_2} \otimes \iota_T), (\mathbf{I}_{N_1} \otimes \iota_{N_2} \otimes \mathbf{I}_T)^*, (\iota_{N_1} \otimes \mathbf{I}_{N_2T})^*);\end{aligned}$$

where $(\cdot)^*$ denotes that the last column of the matrix inside the bracket is deleted. With N_1N_2T rows, they have full column ranks with ranks $N_1 + N_2 + T - 2$, $N_1N_2 + T - 1$, N_1N_2, N_1T, $N_1T + N_2T - 1$ and $N_1N_2 + N_1T + N_2T - 2$ respectively. For different models with particular $\mathbf{D}_\nu$, the construction of π_ν also varies by cases. These explicit expressions of π_ν help readers understand the construction of $\mathbf{D}_\nu$ as they look complicated at the first sight.

$$\begin{aligned}\pi_1 =& (\alpha_1, \cdots, \alpha_{N_1}, \gamma_1, \cdots, \gamma_{N_2-1}, \lambda_1, \cdots, \lambda_{T-1})'_{(N_1+N_2+T-2)\times 1};\\ \pi_2 =& (\gamma_{11}, \cdots, \gamma_{1N_2}, \cdots, \gamma_{N_11}, \cdots, \gamma_{N_1N_2}, \lambda_1, \cdots, \lambda_{T-1})'_{(N_1N_2+T-1)\times 1};\\ \pi_3 =& (\gamma_{11}, \cdots, \gamma_{1N_2}, \cdots, \gamma_{N_11}, \cdots, \gamma_{N_1N_2})'_{N_1N_2\times 1};\end{aligned}$$

$$\pi_4 = (\alpha_{11}, \cdots, \alpha_{1T}, \cdots, \alpha_{N_2 1}, \cdots, \alpha_{N_2 T})'_{N_2 T\times 1};$$
$$\pi_5 = \begin{pmatrix} \alpha_{11}, \cdots, \alpha_{1T}, \cdots, \alpha_{N_2 1}, \cdots, \alpha_{N_2 T} \\ , \alpha^*_{11}, \cdots, \alpha^*_{1T}, \cdots, \alpha^*_{N_1 1}, \cdots, \alpha^*_{N_1 (T-1)} \end{pmatrix}'_{(N_1 T + N_2 T)\times 1};$$
$$\pi_6 = \begin{pmatrix} \gamma_{11}, \cdots, \gamma_{1N_2}, \cdots, \gamma_{N_1 1}, \cdots, \gamma_{N_1 N_2} \\ , \alpha_{11}, \cdots, \alpha_{1T}, \cdots, \alpha_{N_2 1}, \cdots, \alpha_{N_2 (T-1)} \\ , \alpha^*_{11}, \cdots, \alpha^*_{1T}, \cdots, \alpha^*_{N_1 1}, \cdots, \alpha^*_{N_1 (T-1)} \end{pmatrix}'_{(N_1 N_2 + N_1 T + N_2 T - 2)\times 1}.$$

The typical within projector is defined as an idempotent and symmetric matrix

$$\mathbf{M}_{D,\nu} = \mathbf{I} - \mathbf{D}_\nu \left(\mathbf{D}'_\nu \mathbf{D}_\nu\right)^{-1} \mathbf{D}'_\nu.$$

There exist other within projectors which will be discussed in the following sections. Note that some projectors could be nonidempotent, though the projectors we consider in this paper are idempotent. In particular, the linear combination of projectors could be a projector, but it would not always be idempotent. The general model is tranformed to

$$\mathbf{M}_{D,\nu}\mathbf{y} = \rho \mathbf{M}_{D,\nu}\mathbf{y}_{-1} + \mathbf{M}_{D,\nu}\mathbf{X}\beta + \mathbf{M}_{D,\nu}\varepsilon.$$

For two-dimensional panel data models, we can also write down the projectors in this way. For example, consider a PAR-X(1) model with individual fixed effect

$$y_{it} = \rho y_{i,t-1} + \mathbf{x}'_{it}\beta + \alpha_i + \varepsilon_{it}$$

In matrix form,

$$\mathbf{y} = \rho \mathbf{y}_{-1} + \mathbf{X}\beta + \check{\mathbf{D}}_1 \check{\pi}_1 + \varepsilon$$
$$\check{\mathbf{D}}_1 = \mathbf{I}_{N_1} \otimes \iota_T \text{ and } \check{\pi}_1 = (\alpha_1, \cdots, \alpha_{N_1})'$$

The within projector is defined as $\check{\mathbf{M}}_{D,1} = \mathbf{I} - \check{\mathbf{D}}_1 \left(\check{\mathbf{D}}'_1 \check{\mathbf{D}}_1\right)^{-1} \check{\mathbf{D}}'_1$. The transformed dependent variable is $\check{y}_{it,1} = y_{it} - \bar{y}_{i\cdot}$, which eliminates the individual fixed effect component α_i. The fact is that there exist other possible transformations like $\check{y}_{it,2} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot}$, which also eliminates the individual fixed effects. The corresponding within projector for this transformation is $\check{\mathbf{M}}_{D,2} = \mathbf{I} - \check{\mathbf{D}}_2 \left(\check{\mathbf{D}}'_2 \check{\mathbf{D}}_2\right)^{-1} \check{\mathbf{D}}'_2$, where $\check{\mathbf{D}}_2 = \left((\mathbf{I}_{N_1} \otimes \iota_T)^*, \iota_{N_1} \otimes \mathbf{I}_T\right)$ and $(\mathbf{I}_{N_1} \otimes \iota_T)^*$ is the matrix of first $N_1 - 1$ colum vectors. Apparantly it is not a good estimator because it "over-clear" the data in that it makes redundant projections.

To get the estimate of ρ, we take transformation through another idempotent and symmetric matrix $\Xi_X = \mathbf{I} - \mathbf{M}_D\mathbf{X}\left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_D$. The subscript ν is omitted for simplicity. The within estimator of ρ is given by

$$\begin{aligned}\hat{\rho} &= \left(\mathbf{y}_{-1}'\mathbf{M}_D\Xi_X\mathbf{M}_D\mathbf{y}_{-1}\right)^{-1}\mathbf{y}_{-1}'\mathbf{M}_D\Xi_X\mathbf{M}_D\mathbf{y} \\ &= \left(\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1} - \mathbf{y}_{-1}'\mathbf{M}_D\mathbf{X}\left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}_{-1}\right)^{-1} \\ &\quad \times \left(\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y} - \mathbf{y}_{-1}'\mathbf{M}_D\mathbf{X}\left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}\right),\end{aligned} \tag{3.2.1}$$

and the coefficient β of covariates $\mathbf{X}$ is estimated by

$$\begin{aligned}\hat{\beta} &= \left(\mathbf{X}'\mathbf{M}_D\mathbf{X} - \mathbf{X}\mathbf{M}_D\mathbf{y}_{-1}\left(\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right)^{-1}\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{X}\right)^{-1} \\ &\quad \times \left(\mathbf{X}'\mathbf{M}_D\mathbf{y} - \mathbf{X}\mathbf{M}_D\mathbf{y}_{-1}\left(\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right)^{-1}\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}\right).\end{aligned} \tag{3.2.2}$$

For better understanding the notations, we also write them down in scalar forms.

$$\hat{\rho} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} m_{ijt}\left(\ddot{y}_{ij,t-1}^{(-1)}\right)^2\right)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} m_{ijt}\ddot{y}_{ij,t-1}^{(-1)}\ddot{y}_{ijt}$$

$$\text{where } m_{ijt} = 1 - \ddot{\mathbf{x}}_{ijt}'\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{ijt}\ddot{\mathbf{x}}_{ijt}'\right)^{-1}\ddot{\mathbf{x}}_{ijt},$$

and

$$\hat{\beta} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} l_{ijt}\ddot{\mathbf{x}}_{ijt}\ddot{\mathbf{x}}_{ijt}'\right)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} l_{ijt}\ddot{\mathbf{x}}_{ijt}\ddot{y}_{ijt}$$

$$\text{where } l_{ijt} = 1 - \frac{\left(\ddot{y}_{ij,t-1}^{(-1)}\right)^2}{\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\left(\ddot{y}_{ij,t-1}^{(-1)}\right)^2}.$$

where $\ddot{\mathbf{x}}_{ijt}$, $\ddot{y}_{ijt}$ and $\ddot{y}_{ij,t-1}^{(-1)}$ are demeaned observations in which the way of demeaning depends on the within projectors $\mathbf{M}_D$. This estimator could be consistent or inconsistent. The inconsistency comes from the so-called Nickell bias problem. An consistent estimator using Arellano-Bond GMM fashion was proposed by Balazsi et al (2018). In this paper, we suggest another method to handle the Nickell bias, a split-sample jackknife bias correction method (see Dhaene and Jochmans (2015)). Before the asymptotic properties are given, we start with some assumptions. The assumptions follow Moon and Weidner (2017).

Assumption A1: $\{(\ddot{\mathbf{x}}_{ijt}, \varepsilon_{ijt}), t = 1, \ldots, T\}$ are independent across i and j.

Assumption A2: ε_{ijt}'s are independent over t, for all i and j.

Assumption A3: $E[\varepsilon_{ijt}|\sigma\left(\{(\ddot{y}_{ij0}, \ddot{\mathbf{x}}_{ijs}, \varepsilon_{ij,s-1}), s \leq t\}\right)] = 0$ and $E\left(\varepsilon_{ijt}^4\right) < \infty$ for all i, j, and t.

Assumption A4: $\frac{1}{N_1 N_2 T^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{t,s,u,v=1}^{T} |\text{Cov}\left(\varepsilon_{ijt}\ddot{x}_{k,ijs}, \varepsilon_{iju}\ddot{x}_{l,ijv}\right)| = O_p(1)$ for all $k, l = 1, \dots, d$.

Assumption A5: $E\left(\|x_{ijt}\|^4\right) < \infty$ for all i, j, and t.

Assumption A6: $E\left(\pi_{ijt}^4\right) < \infty$ for all i, j, t and $E\left(y_{ij0}^4\right) < \infty$ for all i, j.

If we impose the same fixed effects structure in $\mathbf{x}_{ijt}$ as that in the main equation, it's trivial to fulfill the requirements of A1, A3 and A4, loosely speaking. For example, in Model 1 we can assume that $\mathbf{x}_{ijt} = \tau_i + \varrho_j + \mu_t + \mathbf{e}_{ijt}$, where $\mathbf{e}_{ijt}$ is independent of ε_{ijt}. This specification is similar to the case in Section 5 of Bai (2013). However, if we impose $\mathbf{x}_{ijt} = \tau_{ij} + \mathbf{e}_{ijt}$ and the independence between $\mathbf{e}_{ijt}$ and ε_{ijt}, the inference could be problematic if there exists correlation between τ_{ij} and the fixed effects in the main equation of Model 1. This problem will be revisisted in Section 5. Assumption A1, A2 and A3 are imposed to construct a martingale difference sequence in the proof of consistency of within estimators. Assumption A4, A5 and A6 are used for the proof of asymptotic normality. It's worth noticing that we also put some assmptions on the initial value y_{ij0}. Theorem 1 establishes the asymptotic properties of $\hat{\rho}$, the general within estimators.

Theorem 3.2.1. *Under Assumption A1-A6, the generic estimator $\hat{\rho}$ has the following the asymptotic distribution when N_1, N_2, and T all go to infinity:*

$$\sqrt{N_1 N_2 T}\left(\begin{bmatrix}\hat{\rho}\\ \hat{\beta}\end{bmatrix} - \begin{bmatrix}\rho\\ \beta\end{bmatrix} - \mathbf{\Phi}^{-1}\begin{bmatrix}\frac{1}{N_1 N_2 T} tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\\ \mathbf{0}\end{bmatrix}\right) \to_d N\left(\mathbf{0}, \mathbf{\Phi}^{-1}\mathbf{\Omega}\mathbf{\Phi}^{-1}\right), \tag{3.2.3}$$

where

$$\begin{aligned}
\mathbf{\Phi} &= \begin{bmatrix}\mathbf{A}_\varepsilon & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{A}_\beta & \mathbf{A}'_{x\Lambda}\\ \mathbf{A}_{x\Lambda} & \mathbf{A}_{xx}\end{bmatrix},\\
\mathbf{\Omega} &= \begin{bmatrix}\mathbf{\Omega}_{11} & \mathbf{\Omega}_{12}\\ \mathbf{\Omega}'_{12} & \mathbf{\Omega}_{22}\end{bmatrix},\\
\mathbf{A}_\varepsilon =& \frac{1}{N_1 N_2 T}\left[\frac{1}{1-\rho^2} tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + \frac{2\rho}{1-\rho^2} tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right],\\
\mathbf{A}_\beta =& \frac{1}{N_1 N_2 T}\left[\frac{1}{1-\rho^2}\beta'\mathbf{A}_{xx}\beta + \frac{\rho}{1-\rho^2}\left(\mathbf{A}'_{x\Lambda}\beta + \beta'\mathbf{A}_{x\Lambda}\right)\right],\\
\mathbf{A}_{x\Lambda} =& \frac{1}{N_1 N_2 T} E\left[\mathbf{X}'\mathbf{M}_D\mathbf{\Lambda}\right],
\end{aligned}$$

$$\mathbf{A}_{xx} = \frac{1}{N_1 N_2 T} E\left[\mathbf{X}'\mathbf{M}_D\mathbf{X}\right],$$

$$\mathbf{\Omega}_{11} = \frac{1}{N_1 N_2 T}\left\{\begin{array}{c} E\left(\mathbf{\Lambda}\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{\Lambda}'\right) + E\left(\varepsilon'\mathbf{\Gamma}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{\Gamma}\varepsilon\right) \\ +2E\left(\mathbf{\Lambda}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{\Gamma}\varepsilon\right) - \left[tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2 \end{array}\right\},$$

$$\mathbf{\Omega}_{12} = \frac{1}{N_1 N_2 T}\left[E\left(\mathbf{\Lambda}\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{X}\right) + E\left(\varepsilon\mathbf{\Gamma}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{X}\right)\right],$$

$$\mathbf{\Omega}_{22} = \frac{1}{N_1 N_2 T} E\left(\mathbf{X}'\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{X}\right),$$

$$\mathbf{\Lambda} = \left(\mathbf{I} + \rho\mathbf{B}\right)\left(\mathbf{I} - \rho\mathbf{B}\right)^{-1}\left(\mathbf{y}_0 \otimes \mathbf{e}_1\right) + \mathbf{\Gamma}\mathbf{X}\beta + \mathbf{\Gamma}\mathbf{D}\pi,$$

and $\mathbf{\Sigma}$ is a $N_1N_2T \times N_1N_2T$ diagonal matrix with diagonal item σ^2_{ijt}. Note that the correlation between $\mathbf{X}$ and π enters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ through $\mathbf{\Lambda}$, which indicates the incidental parameter effect. For the simple AR(1) cases, where the general model is given by $\mathbf{y} = \rho\mathbf{y}_{-1} + \mathbf{D}\pi + \varepsilon$, we have the corresponding result

$$\sqrt{N_1 N_2 T}\left(\hat{\rho} - \rho - B_{Nickell}\right) \to_d N\left(0, \Upsilon\right),$$

where

$$B_{Nickell} = \frac{1-\rho^2}{2\rho} p\lim\left(\frac{tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)}{2\rho tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)}\right),$$

$$\Upsilon = \left(\frac{1-\rho^2}{2\rho}\right)^2 p\lim\left(\frac{E\left[\left(\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\right)^2\right] - \left[tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2}{\left[2\rho tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2}\right).$$

The Nickell bias formula $B_{Nickell}$ is the same as equation (33) of Balazsi et al (2018) and the bias expression in Lu et al.(2020, Appendix C).

From Theorem 1, we have that the general fixed effect estimators could be inconsistent. The analysis of asymptotic biases lies in the calculation of $tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)$ and $\mathbf{\Phi}^{-1}$. Different from the traces shown in Table 2 of Balazsi et al (2018), the calculation in our cases are more complicated by introducing heteroskedasticity and the covariates $\mathbf{X}$. The following corollary states the over-clearing facts especially for multilevel panel data models.

Corollary 3.2.1. *For a feasible projector $M_{D,s}$, we can always find another feasible projector $\tilde{M}_{D,s} = M_{D,s}M_{D,t}$ where $M_{D,t}$ is a demean matrix. As defined in Theorem 2.1, denote Ω_s, $\tilde{\Omega}_s$, $\mathbf{\Phi}_s$ and $\tilde{\mathbf{\Phi}}_s$ as the corresponding matrices of the sandwich form in Equation (3.2.3). Therefore, under assumptions of Theorem 2.1, we have $\Omega_s = \tilde{\Omega}_s$ and $\mathbf{\Phi}_s \geq \tilde{\mathbf{\Phi}}_s$. Therefore, the asymptotic variance of the fixed effect estimator of $\tilde{M}_{D,s}$ is larger than that of $\tilde{M}_{D,s}$, i.e., the projector $\tilde{M}_{D,s}$ overclears the data.*

From Corollary 2.1 we have that if we exclude the linear combination of different projectors, some projectors are less efficient in general since they overclear the data. In the following section, we consider the bases for the projectors case by case. Those projectors are competitively efficient and we can construct more competitively efficient fixed effect estimators based on their linear combinations.

3.3 Nickell Bias Representation

3.3.1 Model 1

Model 1 was firstly proposed by Matyas (1997), in which he did not consider the inclusion of X. In order to make the notaions understandable, we consider both the scalar form and the matrix form for the within estimator of model 1. For other five models, we only write them in the matrix form.

There are many different ways to construct $\ddot{y}_{ij,t-1}^{(-1)}$ and $\ddot{y}_{ijt}$ in order to eliminate the fixed effects. To see that, we can find five typical estimators in which $\ddot{y}_{ij,t-1}^{(-1)}$, $\ddot{y}_{ijt}$ and corresponding $\mathbf{M}_{D,1}$ are defined as

a. $\ddot{y}_{ijt,1a} \equiv y_{ijt} - \bar{y}_{\cdot jt} - \bar{y}_{i\cdot t} + \bar{y}_{\cdot\cdot t}$, $\ddot{y}_{ij,t-1,1a}^{(-1)} \equiv y_{ij,t-1} - \bar{y}_{\cdot j,t-1} - \bar{y}_{i\cdot,t-1} + \bar{y}_{\cdot\cdot,t-1}$ and

$$\mathbf{M}_{D,1a} = \mathbf{I} - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2T} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \mathbf{I}_T + \bar{\mathbf{J}}_{N_1N_2} \otimes \mathbf{I}_T$$

b. $\ddot{y}_{ijt,1b} \equiv y_{ijt} - \bar{y}_{i\cdot t} - \bar{y}_{ij\cdot} + \bar{y}_{i\cdot\cdot}$, $\ddot{y}_{ij,t-1,1b}^{(-1)} \equiv y_{ij,t-1} - \bar{y}_{i\cdot,t-1} - \bar{y}_{ij\cdot}^{(-1)} + \bar{y}_{i\cdot\cdot}^{(-1)}$ and

$$\mathbf{M}_{D,1b} = \mathbf{I} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \mathbf{I}_T - \mathbf{I}_{N_1N_2} \otimes \bar{\mathbf{J}}_T + \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2T}$$

c. $\ddot{y}_{ijt,1c} \equiv y_{ijt} - \bar{y}_{\cdot jt} - \bar{y}_{ij\cdot} + \bar{y}_{\cdot j\cdot}$, $\ddot{y}_{ij,t-1,1c}^{(-1)} \equiv y_{ij,t-1} - \bar{y}_{\cdot j,t-1} - \bar{y}_{ij\cdot}^{(-1)} + \bar{y}_{\cdot j\cdot}^{(-1)}$ and

$$\mathbf{M}_{D,1c} = \mathbf{I} - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2T} - \mathbf{I}_{N_1N_2} \otimes \bar{\mathbf{J}}_T + \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T$$

d. $\ddot{y}_{ijt,1d} \equiv y_{ijt} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot t} + 2\bar{y}_{\cdot\cdot\cdot}$, $\ddot{y}_{ij,t-1,1d}^{(-1)} \equiv y_{ij,t-1} - \bar{y}_{i\cdot\cdot}^{(-1)} - \bar{y}_{\cdot j\cdot}^{(-1)} - \bar{y}_{i\cdot,t-1} + 2\bar{y}_{\cdot\cdot\cdot}^{(-1)}$ and

$$\mathbf{M}_{D,1d} = \mathbf{I} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2T} - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_{N_1N_2} \otimes \mathbf{I}_T + \mathbf{2}\bar{\mathbf{J}}_{N_1N_2T}$$

e. $\ddot{y}_{ijt,1e} \equiv y_{ijt} - \bar{y}_{\cdot jt} - \bar{y}_{i\cdot t} - \bar{y}_{ij\cdot} + \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot t} - \bar{y}_{\cdot\cdot\cdot}$, $\ddot{y}_{ij,t-1,1e}^{(-1)} \equiv y_{ij,t-1} - \bar{y}_{\cdot j,t-1} - \bar{y}_{i\cdot,t-1} -$

$\bar{y}_{ij\cdot}^{(-1)} + \bar{y}_{i\cdot\cdot}^{(-1)} + \bar{y}_{\cdot j\cdot}^{(-1)} + \bar{y}_{\cdot\cdot,t-1} - \bar{y}_{\cdots}^{(-1)}$ and

$$\begin{aligned}\mathbf{M}_{D,1e}=&\mathbf{I} - \bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T\\&+\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}+\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T+\bar{\mathbf{J}}_{N_1N_2}\otimes\mathbf{I}_T-\bar{\mathbf{J}}_{N_1N_2T}\end{aligned}$$

Althogh some of them are inconsistent, we can construct five consistent estimators using the split-sample jackknife technique. Without further analysis, we cannot be fully sure that all the possible competitvely efficient estimators have been considered and we cannot fully sure that all the five jackkinfe estimators are good. The Nickell biases can be different. If we fix T and let N_1, N_2 go to infinity, the estimator from (1) is consistent while other four estimators are neither consistent. Also, we find that the combination of some construction can be a new estimator, e.g., $\ddot{y}_{ijt,1e} = \left(\ddot{y}_{ijt,1a} + \ddot{y}_{ijt,1b} + \ddot{y}_{ijt,1c} - \ddot{y}_{ijt,1d}\right)/2$. In the same way, we can contruct lots of (infinite, in fact) estimators like $\ddot{y}_{ijt,1f} = 2 \times \ddot{y}_{ijt,1a} - \ddot{y}_{ijt,1b} = y_{ijt} - 2\bar{y}_{\cdot jt} - \bar{y}_{i\cdot t} + \bar{y}_{ij\cdot} + \bar{y}_{\cdot\cdot t} - \bar{y}_{i\cdot\cdot}$. We deem this phenomeon as the main difference between two-dimensional panel data models and multilevel panel data models. Even if we merely turn our attention from two-dimensional case to three-dimensional case, the model becomes quite complicated. It could be an important question that how to construct and choose a proper fixed effect estimator in high-dimensional panel data models in the future work.

We now discuss the relationship between the fixed effect estimators in the similar patterns.

Proposition 3.3.1. *Under the assumptions of Theorem 2.1, all the within estimators can be generalized based on the first four projectors*

(a) $\mathbf{M}_{D,1a}=\mathbf{I} - \bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T+\bar{\mathbf{J}}_{N_1N_2}\otimes\mathbf{I}_T$;

(b) $\mathbf{M}_{D,1b}=\mathbf{I} - \mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}$;

(c) $\mathbf{M}_{D,1c}=\mathbf{I} - \bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$;

(d) $\mathbf{M}_{D,1d}=\mathbf{I} - \mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T-\bar{\mathbf{J}}_{N_1N_2}\otimes\mathbf{I}_T+\mathbf{2}\bar{\mathbf{J}}_{N_1N_2T}$;

and they are linear independent to each other. In other words, any other within transformation $\ddot{y}_{ijt,1\nu}$ *can be written as the linear combination of* $\ddot{y}_{ijt,1a}$, $\ddot{y}_{ijt,1b}$, $\ddot{y}_{ijt,1c}$ *and* $\ddot{y}_{ijt,1d}$,

$$\ddot{y}_{ijt,1\nu} = \sum_{m\in\{a,b,c,d\}} c_m \ddot{y}_{ijt,1m},\ s.t.,\ \sum_{m=1}^{4} c_m = 1.$$

The proof is straightforward, but the result varies case by case. Before giving the asymptotic distributions, we define $\bar{\sigma} = \frac{1}{N_1N_2T}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T-1}\frac{\sigma_{ijt}^2}{1-\rho}$ and $\bar{\sigma}_\rho = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T-1}\frac{\rho^{T-t}\sigma_{ijt}^2}{1-\rho}$.

Proposition 3.3.2. *Under the assumptions of Theorem 2.1, the Nickell bias terms of the four within estimators in Proposition 3.1 can be represented by*

$$Bias_{1m} := \mathbf{\Phi}_{1m}^{-1}\left[\frac{1}{N_1N_2T}tr\left(\mathbf{\Sigma\Gamma}'\mathbf{M}_D\right),\mathbf{0}\right]' = \left(\frac{\mathbf{B}_{1,1m}}{T}+\frac{\mathbf{B}_{2,1m}}{T^2}\right)h_{1m},$$

for $m \in \{a,b,c,d\}$, $\mathbf{B}_{1,1m} = -\mathbf{\Phi}_{1m}^{-1}\left[\bar{\sigma}/\left(1-\rho\right),\mathbf{0}\right]'$*, and* $\mathbf{B}_{2,1m} = \mathbf{\Phi}_{1m}^{-1}\left[\bar{\sigma}_\rho/\left(1-\rho\right),\mathbf{0}\right]'$*. The scalar* h_{1m} *depends on* N_1 *and* N_2 *such that* $h_{1a}=0$*,* $h_{1b}=1-N_2^{-1}$*,* $h_{1c}=1-N_1^{-1}$*, and* $h_{1d}=N_1^{-1}+N_2^{-1}-(N_1N_2)^{-1}$*. In particular, under the assumption that* $\frac{N_1}{T}\to\delta_1$ *and* $\frac{N_2}{T}\to\delta_2$ *with* N_1*,* N_2*,* $T\to\infty$*, we have the asymptotic distributions for stacked estiamtors* $\hat{\theta}_{1a}$*,* $\hat{\theta}_{1b}$*,* $\hat{\theta}_{1c}$ *and* $\hat{\theta}_{1d}$ *as follows*

$$\begin{aligned}
\sqrt{N_1N_2T}\left(\hat{\theta}_{1a}-\theta\right) &\to_d N\left(\mathbf{0},\mathbf{\Phi}_{1a}^{-1}\mathbf{\Omega}_{1a}\mathbf{\Phi}_{1a}^{-1}\right)\\
\sqrt{N_1N_2T}\left(\hat{\theta}_{1b}-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{1,1b} &\to_d N\left(\mathbf{0},\mathbf{\Phi}_{1b}^{-1}\mathbf{\Omega}_{1b}\mathbf{\Phi}_{1b}^{-1}\right)\\
\sqrt{N_1N_2T}\left(\hat{\theta}_{1c}-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{1,1c} &\to_d N\left(\mathbf{0},\mathbf{\Phi}_{1c}^{-1}\mathbf{\Omega}_{1c}\mathbf{\Phi}_{1c}^{-1}\right)\\
\sqrt{N_1N_2T}\left(\hat{\theta}_{1d}-\theta\right) &\to_d N\left(\mathbf{0},\mathbf{\Phi}_{1d}^{-1}\mathbf{\Omega}_{1d}\mathbf{\Phi}_{1d}^{-1}\right)
\end{aligned}$$

3.3.2 Model 2

We can find two typical estimators in which the corresponding $\mathbf{M}_{D,2}$ are defined as

a. $\mathbf{M}_{D,2a}=\mathbf{I}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}=\mathbf{M}_{D,1b}$;

b. $\mathbf{M}_{D,2b}=\mathbf{I}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T=\mathbf{M}_{D,1c}$;

Proposition 3.3.3. *Under the assumptions of Theorem 2.1, all the within estimators can be generalized based on the two projectors above*

$$\begin{aligned}
\mathbf{M}_{D,2a}&=\mathbf{I}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}\\
\mathbf{M}_{D,2b}&=\mathbf{I}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T+\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T
\end{aligned}$$

and they are linear independent to each other. In other words, any other within transformation $(\nu,\ddot{y}_{ijt})$ *can be written as the linear combination of* $\ddot{y}_{ijt,1a}$ *and* $\ddot{y}_{ijt,1b}$*,*

$$\ddot{y}_{ijt,1\nu}=\sum_{m\in\{a,b\}}c_m\ddot{y}_{ijt,1m},\ s.t.,\ c_1+c_2=1$$

The proof of Proposition 3.2 follows the proof of Proposition 3.1. The Distributions are given in Proposition 3.4.

Proposition 3.3.4. *Under the assumptions of Theorem 2.1, the Nickell bias terms of the two within estimators in Proposition 3.3 can be represented by*

$$\mathbf{Bias}_{2m} := \mathbf{\Phi}_{2m}^{-1}\left[\frac{1}{N_1N_2T}tr\left(\mathbf{\Sigma\Gamma}'\mathbf{M}_{D,2m}\right),\mathbf{0}\right]' = \left(\frac{\mathbf{B}_{1,2m}}{T}+\frac{\mathbf{B}_{2,2m}}{T^2}\right)h_{2m},$$

for $m \in \{a,b,c,d\}$, $\mathbf{B}_{1,2m} = -\mathbf{\Phi}_{2m}^{-1}\left[\bar{\sigma}/\left(1-\rho\right),\mathbf{0}\right]'$, *and* $\mathbf{B}_{2,2m} = \mathbf{\Phi}_{2m}^{-1}\left[\bar{\sigma}_{\rho}/\left(1-\rho\right),\mathbf{0}\right]'$. *The scalar* h_{2m} *depends on* N_1 *and* N_2 *such that* $h_{2a} = 1-N_2^{-1}$ *and* $h_{2b} = 1-N_1^{-1}$. *In particular, under the assumption that* $\frac{N_1}{T}\to\delta_1$ *and* $\frac{N_2}{T}\to\delta_2$ *with* N_1, N_2, $T\to\infty$, *we have the asymptotic distributions for stacked estimators* $\hat{\theta}_{2a}$ *and* $\hat{\theta}_{2b}$ *as follows*

$$\sqrt{N_1N_2T}\left(\hat{\theta}_{2a}-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{2,2a}\to_d N\left(\mathbf{0},\mathbf{\Phi}_{2a}^{-1}\mathbf{\Omega}_{2a}\mathbf{\Phi}_{2a}^{-1}\right)$$
$$\sqrt{N_1N_2T}\left(\hat{\theta}_{2b}-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{2,2b}\to_d N\left(\mathbf{0},\mathbf{\Phi}_{2b}^{-1}\mathbf{\Omega}_{2b}\mathbf{\Phi}_{2b}^{-1}\right)$$

3.3.3 Model 3, Model 4, Model 5, and Model 6

For model 3, 4, 5, and 6, we can find only one good projector for each case.

(1) Model 3: $\mathbf{M}_{D,3}= \mathbf{I}-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T$

(2) Model 4: $\mathbf{M}_{D,4}= \mathbf{I}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}$

(3) Model 5: $\mathbf{M}_{D,5}= \mathbf{I}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T+\bar{\mathbf{J}}_{N_1N_2}\otimes\mathbf{I}_T$

(4) Model 6: $\mathbf{M}_{D,6}= \mathbf{I}-\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2T}-\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T-\mathbf{I}_{N_1N_2}\otimes\bar{\mathbf{J}}_T$
$+\bar{\mathbf{J}}_{N_1N_2}\otimes\mathbf{I}_T+\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T+\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2T}-\bar{\mathbf{J}}_{N_1N_2T}$

There asymptotic properties are given in the following proposition.

Proposition 3.3.5. *Under the assumptions of Theorem 2.1, the Nickell bias terms of the four within estimators above can be represented by (1)* $\mathbf{Bias}_3 = -\frac{\mathbf{B}_{1,3}}{T}+\frac{\mathbf{B}_{2,3}}{T^2}$; *(2)* $\mathbf{Bias}_4 = \mathbf{0}$; *(3)* $\mathbf{Bias}_5 = \mathbf{0}$; *(4)* $\mathbf{Bias}_6 = \left(-\frac{\mathbf{B}_{1,6}}{T}+\frac{\mathbf{B}_{2,6}}{T^2}\right)h_6$ *where* $h_6 = 1-N_2^{-1}-N_1^{-1}+(N_1N_2)^{-1}$, $\mathbf{B}_{1,m} = -\mathbf{\Phi}_m^{-1}\left[\bar{\sigma}/\left(1-\rho\right),\mathbf{0}\right]'$, *and* $\mathbf{B}_{2,m} = \mathbf{\Phi}_m^{-1}\left[\bar{\sigma}_{\rho}/\left(1-\rho\right),\mathbf{0}\right]'$. *In particular, (i) under the assumption that* $\frac{N_1}{T}\to\delta_1$ *and* $\frac{N_2}{T}\to\delta_2$ *with* N_1, N_2, $T\to\infty$, *we have the asymptotic* *distributions for stacked estimators $\hat{\theta}_3$, $\hat{\theta}_4$, $\hat{\theta}_5$ and $\hat{\theta}_6$ as follows*

$$\sqrt{N_1N_2T}\left(\hat{\theta}_3-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{1,3}\rightarrow_d N\left(\mathbf{0},\mathbf{\Phi}_3^{-1}\mathbf{\Omega}_3\mathbf{\Phi}_3^{-1}\right)$$
$$\sqrt{N_1N_2T}\left(\hat{\theta}_4-\theta\right)\rightarrow_d N\left(\mathbf{0},\mathbf{\Phi}_4^{-1}\mathbf{\Omega}_4\mathbf{\Phi}_4^{-1}\right)$$
$$\sqrt{N_1N_2T}\left(\hat{\theta}_5-\theta\right)\rightarrow_d N\left(\mathbf{0},\mathbf{\Phi}_5^{-1}\mathbf{\Omega}_5\mathbf{\Phi}_5^{-1}\right)$$
$$\sqrt{N_1N_2T}\left(\hat{\theta}_6-\theta\right)-\sqrt{\frac{N_1N_2}{T}}\mathbf{B}_{1,6}\rightarrow_d N\left(\mathbf{0},\mathbf{\Phi}_6^{-1}\mathbf{\Omega}_6\mathbf{\Phi}_6^{-1}\right)$$

3.3.4 Comparison with Two-dimensional Panel Data Models

Take Model 3 as an example, the Nickell bias is given by

$$\text{plim}(\hat{\rho}-\rho)=-\frac{1}{T}\mathbf{\Phi}_3^{-1}\frac{\bar{\sigma}}{1-\rho}+\frac{1}{T^2}\mathbf{\Phi}_3^{-1}\frac{\bar{\sigma}_\rho}{1-\rho}.$$

As shown in Nickell (1981), for a two-dimensional panel data model, $y_{it}=\rho y_{it-1}+\beta'\mathbf{x}_{it}+\gamma_i+\varepsilon_{it}$, the bias is given by

$$\text{plim}(\hat{\rho}-\rho)=-\frac{1}{T}\mathbf{\Phi}_{Nickell}^{-1}\frac{\bar{\sigma}_{Nickell}}{1-\rho}+\frac{1}{T^2}\mathbf{\Phi}_{Nickell}^{-1}\frac{\bar{\sigma}_{\rho,Nickell}}{1-\rho},$$

where $\mathbf{\Phi}_{Nickell}=\text{plim}\frac{1}{N}\tilde{y}'M\tilde{y}$, $\tilde{y}_t=y_{it}-y_{i\cdot}$, $M=I-\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$, $\tilde{X}_t=x_{ijt}-x_{ij\cdot}$, $\bar{\sigma}_{Nickell}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\frac{\sigma_{it}^2}{1-\rho}$ and $\bar{\sigma}_\rho=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\frac{\rho^{T-t}\sigma_{it}^2}{1-\rho}$. We can see that they have similar structure.

3.4 Split-panel Jackknife Estimation

Based on the results of Proposition 3.2, 3.4 and 3.5, we can construct the split-panel jackknife estimator. As said in Dhaene and Jochmans (2015), over-spliting the sample will increase the magnitude of higher-order bias terms. Therefore, we only consider "half-panel" jackknife estimators in this paper. We define $S_1=\{S_{11},S_{12}\}$, where $S_{11}:=\{1,2,\ldots,\lceil T/2\rceil\}$ and $S_{12}:=\{\lceil T/2\rceil+1,\ldots,T\}$ splitting over dimension t; $S_2=\{S_{21},S_{22}\}$, where $S_{21}:=\{1,2,\ldots,\lfloor T/2\rfloor\rceil\}$ and $S_{12}:=\{\lfloor T/2\rfloor+1,\ldots,T\}$ splitting over dimension t. The split-panel jackknife estimator is defined as $\tilde{\theta}_{1/2}=2\hat{\theta}-\frac{1}{2}\left(\overline{\theta}_{S_1}+\overline{\theta}_{S_2}\right)$, where $\overline{\theta}_{S_m}:=\sum_{S_{mk}\in S_m}\frac{|S_{mk}|}{T}\hat{\theta}_{S_{mk}}$.

Note that when T is even, $\lceil T/2\rceil=\lceil T/2\rceil$, thus $S_1=S_2$ and $\tilde{\theta}_{1/2}=2\hat{\theta}-\overline{\theta}_{S_1}$. Similar to Theorem 3.1 of Dhaene and Jochmans (2015), we have the following Theorem 2.

Theorem 3.4.1. *Under Assumption 1-4, the general estimator $\hat{\rho}$ has the following the asymptotic distribution when $\frac{N_1}{T} \to \delta_1$ and $\frac{N_2}{T} \to \delta_2$ with N_1, N_2, $T \to \infty$,*

$$\sqrt{N_1 N_2 T}\left(\tilde{\theta}_{1/2} - \theta\right) \to_d N\left(\mathbf{0}, \mathbf{\Phi}^{-1}\mathbf{\Omega}\mathbf{\Phi}^{-1}\right) \tag{3.4.1}$$

Note that the correlation between $\mathbf{X}$ and π enters $\mathbf{\Phi}$ and $\mathbf{\Omega}$ through $\mathbf{\Lambda}$, which indicates the incidental parameter effect. For the simple AR(1) cases, where the general model is given by $\mathbf{y} = \rho\mathbf{y}_{-1} + \mathbf{D}\pi + \varepsilon$, we have the corresponding result

$$\sqrt{N_1 N_2 T}\left(\tilde{\rho}_{1/2} - \rho\right) \to_d N\left(0, \Upsilon\right)$$

where

$$\Upsilon = \left(\frac{1-\rho^2}{2\rho}\right)^2 p\lim\left(\frac{E\left[\left(\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\right)^2\right] - \left[tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2}{\left[2\rho tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2}\right)$$

The asymptotic biases is eliminated by the split-panel jackknife estimators.

3.5 Model Uncertainty

In the simplest fixed effect panel data models with other covariates, we need to impose the assumption of IID on incidental parameters in order to get efficient within estimators. If we consider a more complicated model, a two-way error component panel data model, the IID assumption of α_i and λ_t is not enough. The reason is that the correlation between individual invariant term λ_t and covariates x_{it} will leave an unspecified dependence structure on x_{it} and x_{jt}. For this perspective, we need to impose structure on x_{it}. As in Bai (2013), they assumed that $x_{it} = \tau_i + b_t + \rho_x x_{i,t-1} + e_{it}$ (see p. 298 in Bai (2013)) where e_{it} is independent with the idiosyncratic error term in the equation of y_{it}. This problem also exists in more general cases. For the panel data models with interactive fixed effects, we also need to impose some structures on x_{it}. As in Moon and Weidner (2017), for the cross section dimension they assume that $\{X_{it}\}$ is independent across i conditional on $\mathcal{C}$, the sigma-algebra generated by the factors and the factor loadings (see Assumption 5(iii) in Moon and Weidner (2015)); for the time dimension they assumed that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s,u,v=1}^{T}\left|\text{Cov}\left(\varepsilon_{it}\tilde{X}_{k,is}, \varepsilon_{iu}\tilde{X}_{l,iv}|\mathcal{C}\right)\right| = O_p(1)$, where $\tilde{X}_{k,it} = X_{k,it} - E\left[X_{k,it}|\mathcal{C}\right]$ is the projection residual of $X_{k,it}$ after being projected through $\mathcal{C}$ (see Assumption 5(v) in Moon and Weidner (2017)).

Even though we exclude the cross-sectional dependence in this paper, in the multilevel panel data models, we have seen that there exist many possible fixed effects structures in the expression equation for y_{it}, which is similar to the cases in Moon and Weidner (2017).

Questions are raised that does the misspecification of x_{ijt} affects the inference on the within estimators and if does how shall we handle that. We deem this kind of misspecification as the model uncertainty lying in the nature of multilevel panel data models.

3.5.1 Focused Model Selection

In this section, we will mainly consider the selection problem in Model 1 and Model 2. Taking Model 2 as an example, the performance of $\hat{\theta}_{2a}$ and $\hat{\theta}_{2b}$ depends on the specification of $\mathbf{x}_{ijt}$ even when we assume Model 2 is true. If $\mathbf{x}_{ijt}$ has different structure as $\tilde{\gamma}_{ij} + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}$, we will see that different projectors will eliminate different variations in $\mathbf{x}_{ijt}$. There are two implication for this statement. The first is that if $\mathbf{x}_{ijt}$ contains some observable variables that does not change over j but change over i and t, i.e., $x_{k,it}$, those variables will vanish after transformed by the first projector but will remian after tranformed by the second projector. The second implication is that if $\mathbf{x}_{ijt}$ has the factor structure like $\tilde{\alpha}_{it} + \tilde{\gamma}_{ij} + \tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}$, $\tilde{\gamma}_{ij}$ and $\tilde{\lambda}_t$ will vanish for both projectors, and $\tilde{\alpha}_{it}$ will be retained only for the second projector. In this sense, the first projector loses more variation in $\mathbf{x}_{ijt}$ in contrast to the first projector. On the contrary, there exist some circumstances that the second projector loses more variation. For the first implication, we suggest that researchers should choose estimators based on the characteristics of variables. However, sometimes we have both $x_{1,it}$ and $x_{2,jt}$ in the regression equation and sometimes we are not able to observe the factor structure of covariates, then we face the problem of model selection.

Claeskens and Hjort (2003) proposed focused information criterion for cross section models with finite-dimensional nuisance parameters. Lee and Phillips (2015) (LP15 hereafter) stated that the reason why traditional model selection method performs bad in panel data models is that the dimension of nuisance parameters grows with the sample size. LP15 also considered the lag order selction problem in dynamic panel data models.

3.6 Conclusion

We consider the Nickell bias problem in dynamic fixed effects multilevel panel data models with various kinds of multi-way error components. For some specifications of error components, there exist many different forms of within estimators which are shown to be of possibly different asymptotic properties. The forms of the estimators in our framework are given explicitly. We apply the split-sample jackknife approach to eliminate the bias. In practice, our results can be easily extended to multilevel panel data models with higher dimensions.

3.7 Appendix

3.7.1 Proof of Theorem 3.2.1

Proof. For clearance, we use bold symbols to represent vectors and matrices. The general dynamic multilevel model we consider is

$$\mathbf{y} = \rho \mathbf{y}_{-1} + \mathbf{X}\beta + \mathbf{D}\pi + \varepsilon$$

Assume that $\mathbf{M}_D$ is a appropriate fixed effects projector, which might not be idempotent. We have the following transformed equation

$$\mathbf{M}_D\mathbf{y} = \rho \mathbf{M}_D\mathbf{y}_{-1} + \mathbf{M}_D\mathbf{X}\beta + \mathbf{M}_D\varepsilon$$

Writing it in the partitioned matrix form, we have

$$\mathbf{M}_D\mathbf{y} = \begin{bmatrix} \mathbf{M}_D\mathbf{y}_{-1} & \mathbf{M}_D\mathbf{X} \end{bmatrix} \begin{bmatrix} \rho \\ \beta \end{bmatrix} + \mathbf{M}_D\varepsilon$$

The generic within estimators for $\begin{bmatrix} \rho \\ \beta \end{bmatrix}$ are defined as

$$\begin{aligned}
\begin{bmatrix} \hat{\rho} \\ \hat{\beta} \end{bmatrix} &= \left[\begin{pmatrix} \mathbf{y}'_{-1}\mathbf{M}_D \\ \mathbf{X}'\mathbf{M}_D \end{pmatrix} \begin{pmatrix} \mathbf{M}_D\mathbf{y}_{-1} & \mathbf{M}_D\mathbf{X} \end{pmatrix} \right]^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{X} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{X}'\mathbf{M}_D\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{X} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{X}'\mathbf{M}_D\mathbf{X} \end{bmatrix}^{-1} \left[\begin{pmatrix} \mathbf{y}'_{-1}\mathbf{M}_D \\ \mathbf{X}'\mathbf{M}_D \end{pmatrix} \begin{pmatrix} \mathbf{M}_D\mathbf{y}_{-1} & \mathbf{M}_D\mathbf{X} \end{pmatrix} \right] \begin{bmatrix} \rho \\ \beta \end{bmatrix} \\
&\quad + \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{X} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{X}'\mathbf{M}_D\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\varepsilon \\ \mathbf{X}'\mathbf{M}_D\varepsilon \end{bmatrix} \\
&= \begin{bmatrix} \rho \\ \beta \end{bmatrix} + \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{y}'_{-1}\mathbf{M}_D\mathbf{X} \\ \mathbf{X}'\mathbf{M}_D\mathbf{y}_{-1} & \mathbf{X}'\mathbf{M}_D\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{M}_D\varepsilon \\ \mathbf{X}'\mathbf{M}_D\varepsilon \end{bmatrix} \\
&= \begin{bmatrix} \rho \\ \beta \end{bmatrix} + \mathbf{A}^{-1}_{N_1N_2T}\mathbf{B}_{N_1N_2T}
\end{aligned}$$

where

$$\mathbf{A}_{N_1N_2T}=\begin{bmatrix}\mathbf{y}_{-1}'\mathbf{M}_D^2\mathbf{y}_{-1} & \mathbf{y}_{-1}'\mathbf{M}_D^2\mathbf{X}\\ \mathbf{X}'\mathbf{M}_D^2\mathbf{y}_{-1} & \mathbf{X}'\mathbf{M}_D^2\mathbf{X}\end{bmatrix} \text{ and } \mathbf{B}_{N_1N_2T}=\begin{bmatrix}\mathbf{y}_{-1}'\mathbf{M}_D^2\varepsilon\\ \mathbf{X}'\mathbf{M}_D^2\varepsilon\end{bmatrix}$$

Firstly we consider the asymptotic limit of $\mathbf{B}_{N_1N_2T}$ when N_1, N_2, and T all goes to infinity. Given the $(N_1N_2T\times 1)$ sample realization vector of dependent variables $\mathbf{y}$, we have that $\mathbf{y}_{-1}=\mathbf{\Psi}\mathbf{y}+\mathbf{y}_0\otimes\mathbf{e}_1$. From the dynamic structure that $\mathbf{y}=\rho\mathbf{y}_{-1}+\mathbf{X}\beta+\mathbf{D}\pi+\varepsilon$, we have $\mathbf{y}=\rho\mathbf{\Psi}\mathbf{y}+\rho\mathbf{y}_0\otimes\mathbf{e}_1+\mathbf{X}\beta+\mathbf{D}\pi+\varepsilon$. Hence $(\mathbf{I}-\rho\mathbf{\Psi})\,\mathbf{y}=\rho\mathbf{y}_0\otimes\mathbf{e}_1+\mathbf{X}\beta+\mathbf{D}\pi+\varepsilon$. Therefore, we have the representation for $\mathbf{y}$.

$$\mathbf{y}=(\mathbf{I}-\rho\mathbf{\Psi})^{-1}(\rho\mathbf{y}_0\otimes\mathbf{e}_1)+(\mathbf{I}-\rho\mathbf{\Psi})^{-1}\mathbf{X}\beta+(\mathbf{I}-\rho\mathbf{\Psi})^{-1}\mathbf{D}\pi+(\mathbf{I}-\rho\mathbf{\Psi})^{-1}\varepsilon$$

and since $\mathbf{y}_{-1}=\mathbf{\Psi}\mathbf{y}+\mathbf{y}_0\otimes\mathbf{e}_1$, we have

$$\mathbf{y}_{-1}=(\mathbf{I}+\rho\mathbf{\Psi})(\mathbf{I}-\rho\mathbf{\Psi})^{-1}(\mathbf{y}_0\otimes\mathbf{e}_1)+\mathbf{\Gamma}\mathbf{X}\beta+\mathbf{\Gamma}\mathbf{D}\pi+\mathbf{\Gamma}\varepsilon:=\mathbf{\Lambda}+\mathbf{\Gamma}\varepsilon$$

where $\mathbf{\Gamma}:=\mathbf{\Psi}(\mathbf{I}-\rho\mathbf{\Psi})^{-1}$ and $\mathbf{\Lambda}:=(\mathbf{I}+\rho\mathbf{\Psi})(\mathbf{I}-\rho\mathbf{\Psi})^{-1}(\mathbf{y}_0\otimes\mathbf{e}_1)+\mathbf{\Gamma}\mathbf{X}\beta+\mathbf{\Gamma}\mathbf{D}\pi$. Denote $\mathbf{Z}=\begin{bmatrix}\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon\\ \mathbf{X}'\mathbf{M}_D\varepsilon\end{bmatrix}$, we have $E(\mathbf{Z})=\begin{bmatrix}E\left(\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon\right)\\ 0\end{bmatrix}$ and

$$\begin{aligned}E\left(\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon\right)=&E\left(\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\right)=E\left(tr\left(\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\right)\right)=E\left(tr\left(\varepsilon\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\right)\right)\\=&tr\left(E\left(\varepsilon\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\right)\right)=tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right).\end{aligned}$$

Therefore $E(\mathbf{Z})=\begin{bmatrix}tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\\ \mathbf{0}\end{bmatrix}$. Let $\mathbf{c}$ be a $(1+d)$-vector such that $\|\mathbf{c}\|=1$. We follow the the Cramer-Wold device and CLT for martingale difference sequences (see Corollary 5.26 in White (1984)) to proceed the proof. The purpose is to verify

$$\frac{1}{\sqrt{N_1N_2T}}\left(\mathbf{c}'\mathbf{Z}-\mathbf{c}'E(\mathbf{Z})\right)\to_d N\left(0,\mathbf{c}'\mathbf{\Omega}\mathbf{c}\right)$$

More precisely,

$$\begin{aligned}\frac{1}{\sqrt{N_1N_2T}}\mathbf{c}'\left[\mathbf{Z}-E(\mathbf{Z})\right]=&\frac{1}{\sqrt{N_1N_2T}}\left(c_0\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon+\tilde{\mathbf{c}}'\mathbf{X}'\mathbf{M}_\mathbf{D}\varepsilon-c_0\varepsilon'\mathbf{\Gamma}'\mathbf{M}_\mathbf{D}\varepsilon\right)\\=&\frac{1}{\sqrt{N_1N_2T}}\left[c_0\mathbf{\Lambda}'\mathbf{M}_D\varepsilon+\tilde{\mathbf{c}}'\mathbf{X}'\mathbf{M}_D\varepsilon\right]\\=&\frac{1}{\sqrt{N_1N_2T}}\mathbf{b}'\mathbf{M}_D\varepsilon\end{aligned}$$

where $\mathbf{b} = c_0\mathbf{\Lambda} + \mathbf{X}\tilde{\mathbf{c}} = c_0\left(\mathbf{I}+\rho\mathbf{\Psi}\right)\left(\mathbf{I}-\rho\mathbf{\Psi}\right)^{-1}\left(\mathbf{y}_0\otimes\mathbf{e}_1\right)+c_0\mathbf{\Gamma}\mathbf{X}\beta+c_0\mathbf{\Gamma}\mathbf{D}\pi+\mathbf{X}\tilde{\mathbf{c}} = \mathbf{G}\left(\mathbf{y}_0,\mathbf{X},\pi\right)$. To prove that, we apply CLT for elements of $\mathbf{cZ}$. In scalar form,

$$\begin{aligned}\frac{1}{\sqrt{N_1N_2T}}\mathbf{b}'\mathbf{M}_D\varepsilon &= \frac{1}{\sqrt{N_1N_2T}}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\left[\ddot{\varepsilon}_{ijt}\left(y_{ij0}+\sum_{s=1}^{t}\rho^{s-1}\left(\rho y_{ij0}+x'_{ijs}\beta\right)c_0+x'_{ijt}\tilde{c}\right)\right]\\ &= \frac{1}{\sqrt{N_1N_2T}}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T}\left[\varepsilon_{ijt}\left(\ddot{y}_{ij0}+\sum_{s=1}^{t}\rho^{s-1}\left(\rho\ddot{y}_{ij0}+\ddot{x}'_{ijs}\beta\right)c_0+\ddot{x}'_{ijt}\tilde{c}\right)\right]\end{aligned}$$

where $\ddot{x}_{ijt}$ denotes the x_{ijt} transformed by the projector $\mathbf{M}_D$. Denote

$$\xi_{ijt} = \varepsilon_{ijt}\left(\ddot{y}_{ij0}+\sum_{s=1}^{t-1}\rho^{t-s}\left(\rho\ddot{y}_{ij0}+\ddot{x}'_{ijt-s}\beta\right)c_0+\ddot{x}'_{ijt}\tilde{c}\right)$$

From Assumption 1, 2 and 3, we have

$$E\left(\xi_{ijt}|\sigma\left(\left\{\left(\ddot{y}_{ij0},\ddot{\mathbf{x}}_{ijs},\varepsilon_{ij,s-1}\right),s\leq t\right\}\right)\right)=0$$

and

$$E\left(\xi_{ijt}^2|\sigma\left(\left\{\left(\ddot{y}_{ij0},\ddot{\mathbf{x}}_{ijs},\varepsilon_{ij,s-1}\right),s\leq t\right\}\right)\right)<\infty.$$

It is shown that ξ_m is a martingale difference sequence. Therefore, applying CLT for martingale difference sequences, $\frac{1}{\sqrt{N_1N_2T}}\mathbf{b}'\mathbf{M}_D\varepsilon \to_d N\left(0,\mathbf{c}'\mathbf{\Omega}\mathbf{c}\right)$ where $\mathbf{\Omega} = \frac{1}{N_{11}N_2T}Var\left(\mathbf{Z}\right) := \frac{1}{N_{11}N_2T}\begin{bmatrix}\mathbf{\Omega}_{11} & \mathbf{\Omega}_{12}\\ \mathbf{\Omega}'_{12} & \mathbf{\Omega}_{22}\end{bmatrix}$. Then we have

$$\begin{aligned}\mathbf{\Omega}_{11} &= E\left(\mathbf{y}'_{-1}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{y}_{-1}\right)-\left[tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2\\ &= E\left(\mathbf{\Lambda}\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{\Lambda}'\right)+E\left(\varepsilon'\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{\Gamma}\varepsilon\right)+2E\left(\mathbf{\Lambda}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{\Gamma}\varepsilon\right)-\left[tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)\right]^2\\ \mathbf{\Omega}_{12} &= E\left(\mathbf{y}'_{-1}\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{X}\right)=E\left(\mathbf{\Lambda}\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{X}\right)+E\left(\varepsilon\mathbf{\Gamma}'\mathbf{M}_D\varepsilon\varepsilon'\mathbf{M}_D\mathbf{X}\right)\\ \mathbf{\Omega}_{22} &= E\left(\mathbf{X}'\mathbf{M}_D\mathbf{\Sigma}\mathbf{M}_D\mathbf{X}\right)\end{aligned}$$

Then we consider the asymptotic limit of $\mathbf{A}_{N_1N_2T}$ when N_1, N_2, and T all goes to infinity. By the weak law of large numbers,

$$\mathbf{A}_{N_1N_2T}\to_p\begin{bmatrix}E\left[\mathbf{y}'_{-1}\mathbf{M}_D^2\mathbf{y}_{-1}\right] & E\left[\mathbf{y}'_{-1}\mathbf{M}_D^2\mathbf{X}\right]\\ E\left[\mathbf{X}'\mathbf{M}_D^2\mathbf{y}_{-1}\right] & E\left[\mathbf{X}'\mathbf{M}_D^2\mathbf{X}\right]\end{bmatrix}$$

For $E\left[\mathbf{y}'_{-1}\mathbf{M}_D\mathbf{y}_{-1}\right]$, we can use the dynamic equation and the stationarity condition to get the result.

$$
\begin{aligned}
E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right] =& E\left[\mathbf{y}'\mathbf{M}_D\mathbf{y}\right] = E\left[\left(\rho\mathbf{y}_{-1}'+\beta'\mathbf{X}'+\varepsilon'\right)\mathbf{M}_D\left(\rho\mathbf{y}_{-1}+\mathbf{X}\beta+\varepsilon\right)\right] \\
=&\rho^2 E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right] + 2\rho E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{X}\beta\right] + E\left[\varepsilon'\mathbf{M}_D\varepsilon\right] \\
&+ E\left[\beta'\mathbf{X}'\mathbf{M}_D\mathbf{X}'\beta\right] + 2\rho E\left[\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon\right]
\end{aligned}
$$

Then we have the expression for $E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right]$,

$$
\begin{aligned}
E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{y}_{-1}\right] =& \frac{1}{1-\rho^2}\left\{tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + E\left[\beta'\mathbf{X}'\mathbf{M}_D\mathbf{X}\beta\right]\right\} \\
&+ \frac{2\rho}{1-\rho^2}\left\{tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right) + E\left[\mathbf{y}_{-1}'\mathbf{M}_D\mathbf{X}\beta\right]\right\}
\end{aligned}
$$

with $E\left[\mathbf{y}_{-1}'\mathbf{M}_D\varepsilon\right] = tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)$ and $E\left[\varepsilon'\mathbf{M}_D\varepsilon\right] = tr\left(\mathbf{\Sigma}\mathbf{M}_D\right)$. To illustrate the limit of $\mathbf{A}_{N_1N_2T}$, it can be shown that

$$
p\lim \frac{1}{N_1N_2T}\mathbf{A}_{N_1N_2T} = \frac{1}{N_1N_2T}\left\{\begin{bmatrix}\mathbf{A}_\varepsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{A}_\beta & \mathbf{A}_{x\Lambda}' \\ \mathbf{A}_{x\Lambda} & \mathbf{A}_{xx}\end{bmatrix}\right\} := \mathbf{\Phi}
$$

where $\mathbf{A}_\varepsilon = \frac{1}{1-\rho^2}tr\left(\mathbf{\Sigma}\mathbf{M}_D\right) + \frac{2\rho}{1-\rho^2}tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)$, $\mathbf{A}_\beta = \frac{1}{1-\rho^2}\beta'\mathbf{A}_{xx}\beta + \frac{\rho}{1-\rho^2}\left(\mathbf{A}_{x\Lambda}'\beta + \beta'\mathbf{A}_{x\Lambda}\right)$, $\mathbf{A}_{x\Lambda} = E\left[\mathbf{X}'\mathbf{M}_D\mathbf{\Lambda}\right]$ and $\mathbf{A}_{xx} = E\left[\mathbf{X}'\mathbf{M}_D\mathbf{X}\right]$.

By Delta method, we have the asymptotic distribution of $\begin{bmatrix}\hat{\rho} \\ \hat{\beta}\end{bmatrix}$,

$$
\sqrt{N_1N_2T}\left(\begin{bmatrix}\hat{\rho} \\ \hat{\beta}\end{bmatrix} - \begin{bmatrix}\rho \\ \beta\end{bmatrix} - \mathbf{\Phi}^{-1}\begin{bmatrix}\frac{tr\left(\mathbf{\Sigma}\mathbf{\Gamma}'\mathbf{M}_D\right)}{N_1N_2T} \\ \mathbf{0}\end{bmatrix}\right) \to_d N\left(\mathbf{0}, \mathbf{\Phi}^{-1}\mathbf{\Omega}\mathbf{\Phi}^{-1}\right)
$$

which completes the proof. □

3.7.2 Proof of Corollary 3.2.1

Proof. Firstly we prove that $\mathbf{\Omega}_s \le \tilde{\mathbf{\Omega}}_s$. Denote $\tilde{\mathbf{Z}} = \left[\mathbf{y}_{-1}', \mathbf{X}'\right]'/\sqrt{N_1N_2T}$ and $\check{\mathbf{Z}} = \left[\mathbf{\Lambda}, \mathbf{X}'\right]'/\sqrt{N_1N_2T}$, we have $\mathbf{\Omega}_s = Var\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\varepsilon\right)$ and $\tilde{\mathbf{\Omega}}_s = Var\left(\tilde{\mathbf{Z}}\tilde{\mathbf{M}}_{D,s}\varepsilon\right) = Var\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\right)$. Since

$$
\begin{aligned}
\mathbf{\Omega}_s - \tilde{\mathbf{\Omega}}_s =& Var\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\varepsilon\right) - Var\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\right) \\
=& E\left[\left(\tilde{\mathbf{Z}} - E\tilde{\mathbf{Z}}\right)\mathbf{M}_{D,s}\varepsilon\right]\left[\left(\tilde{\mathbf{Z}} - E\tilde{\mathbf{Z}}\right)\mathbf{M}_{D,s}\varepsilon\right]' \\
&- E\left[\left(\tilde{\mathbf{Z}} - E\tilde{\mathbf{Z}}\right)\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\right]\left[\left(\tilde{\mathbf{Z}} - E\tilde{\mathbf{Z}}\right)\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\right]'
\end{aligned}
$$

$$
\begin{aligned}
=&E\left[\breve{\mathbf{Z}}\mathbf{M}_{D,s}\varepsilon\varepsilon'\mathbf{M}_{D,s}\breve{\mathbf{Z}}'\right]-E\left[\breve{\mathbf{Z}}\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\varepsilon'\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\breve{\mathbf{Z}}'\right]\\
=&E\left[\breve{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{M}_{D,s}\varepsilon\varepsilon'\mathbf{M}_{D,s}-\mathbf{M}_{D,t}\mathbf{M}_{D,s}\varepsilon\varepsilon'\mathbf{M}_{D,s}\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}\breve{\mathbf{Z}}'\right]\\
=&E\left[\breve{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{M}_{D,s}\varepsilon\varepsilon'\mathbf{M}_{D,s}-\mathbf{M}_{D,s}\mathbf{M}_{D,t}\varepsilon\varepsilon'\mathbf{M}_{D,t}\mathbf{M}_{D,s}\right)\mathbf{M}_{D,s}\breve{\mathbf{Z}}'\right]\\
=&E\left[\breve{\mathbf{Z}}\mathbf{M}_{D,s}\left(\varepsilon\varepsilon'-\mathbf{M}_{D,t}\varepsilon\varepsilon'\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}\breve{\mathbf{Z}}'\right]\\
=&E\breve{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{\Sigma}-\mathbf{M}_{D,t}\mathbf{\Sigma}\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}E\breve{\mathbf{Z}}'
\end{aligned}
$$

Since $\mathbf{\Sigma}$ is a diagonal matrix with positive diagonal item σ^2_{ijt} and $\mathbf{M}_{D,t}\mathbf{\Sigma}\mathbf{M}_{D,t}$ is a diagonal matrix with positive diagonal item $\ddot{\sigma}^2_{ijt}=\sigma^2_{ijt}+o\left(1\right)$, which is the demean version of σ^2_{ijt}, we have that $\mathbf{\Sigma}-\mathbf{M}_{D,t}\mathbf{\Sigma}\mathbf{M}_{D,t}=o(1)$ and then $E\breve{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{\Sigma}-\mathbf{M}_{D,t}\mathbf{\Sigma}\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}E\breve{\mathbf{Z}}'\to 0$ as N_1, N_2 and T tend to infinity. Thus $\mathbf{\Omega}_s=\tilde{\mathbf{\Omega}}_s$.

Secondly we prove that $\mathbf{\Phi}_s\geq\tilde{\mathbf{\Phi}}_s$.

$$
\begin{aligned}
\mathbf{\Phi}_s-\tilde{\mathbf{\Phi}}_s=&E\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\tilde{\mathbf{Z}}'\right)-E\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\mathbf{M}_{D,t}\mathbf{M}_{D,s}\tilde{\mathbf{Z}}'\right)\\
=&E\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{I}-\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}\tilde{\mathbf{Z}}'\right)\\
=&E\left(\tilde{\mathbf{Z}}\mathbf{M}_{D,s}\left(\mathbf{I}-\mathbf{M}_{D,t}\right)\left(\mathbf{I}-\mathbf{M}_{D,t}\right)\mathbf{M}_{D,s}\tilde{\mathbf{Z}}'\right)\geq 0
\end{aligned}
$$

The last equality is obtained from the fact that $\mathbf{I}-\mathbf{M}_{D,t}$ is a symmetric and idempotent matrix. This completes the proof. □

3.7.3 Proof of Proposition 3.3.1

Proof. There are seven kinds of taking averaging in the three-dimensional panel data models: (1) over i, (2) over j, (3) over t, (4) over i and j, (5) over i and t, (6) over j and t, and (7) over i, j, and t. The matrix form and the corresponding transformed incidental parameters are

No.	Scalar Form	Matrix Form	After Trans.	$\mathbf{Q}_m$
0	y_{ijt}	$\mathbf{I}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\mathbf{I}_T$	$\alpha_i+\gamma_j+\lambda_t$	$[1,0,1,0,1,0]$
1	$\bar{y}_{i\cdot t}$	$\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T$	$\alpha_i+\bar{\gamma}+\lambda_t$	$[1,0,0,1,1,0]$
2	$\bar{y}_{ij\cdot}$	$\mathbf{I}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$	$\alpha_i+\gamma_j+\bar{\lambda}$	$[1,0,1,0,0,1]$
3	$\bar{y}_{\cdot jt}$	$\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\mathbf{I}_T$	$\bar{\alpha}+\gamma_j+\lambda_t$	$[0,1,1,0,1,0]$
4	$\bar{y}_{i\cdot\cdot}$	$\mathbf{I}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\bar{\mathbf{J}}_T$	$\alpha_i+\bar{\gamma}+\bar{\lambda}$	$[1,0,0,1,0,1]$
5	$\bar{y}_{\cdot j\cdot}$	$\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$	$\bar{\alpha}+\gamma_j+\bar{\lambda}$	$[0,1,1,0,0,1]$
6	$\bar{y}_{\cdot\cdot t}$	$\bar{\mathbf{J}}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\mathbf{I}_T$	$\bar{\alpha}+\bar{\gamma}+\lambda_t$	$[0,1,0,1,1,0]$
7	$\bar{y}_{\cdot\cdot\cdot}$	$\bar{\mathbf{J}}_{N_1}\otimes\bar{\mathbf{J}}_{N_2}\otimes\bar{\mathbf{J}}_T$	$\bar{\alpha}+\bar{\gamma}+\bar{\lambda}$	$[0,1,0,1,0,1]$

Define the elements of transformed incidental parameters $\eta=\left[\alpha_i,\bar{\alpha},\gamma_j,\bar{\gamma},\lambda_t,\bar{\lambda}\right]$, the transformed incidental parameters can be represented by $\mathbf{Q}_m\eta$ in each averaging case. In the most general case, $\eta=\left[\alpha_i,\bar{\alpha},\gamma_j,\bar{\gamma},\lambda_t,\bar{\lambda},\gamma_{ij},\bar{\gamma}_{\cdot j},\bar{\gamma}_{i\cdot},\bar{\gamma}_{\cdot\cdot},\alpha_{it},\bar{\alpha}_{\cdot t},\bar{\alpha}_{i\cdot},\bar{\alpha}_{\cdot\cdot},\alpha^*_{jt},\bar{\alpha}^*_{\cdot t},\bar{\alpha}^*_{j\cdot},\bar{\alpha}^*_{\cdot\cdot}\right]$. For simplicity of notations, we adopt the special specification $\eta=\left[\alpha_i,\bar{\alpha},\gamma_j,\bar{\gamma},\lambda_t,\bar{\lambda}\right]$ in this proof. The original incidental parameters are $\alpha_i+\gamma_j+\lambda_t$. The purpose of projection is to eliminate them with the linear combinations of these seven kinds of transformations. Since the previous table include all the possible averaging schemes, we can imagine that the potential feasible projection could be equal to the linear combinations of $\mathbf{Q}_m\eta$. In other words, we need to solve the following multivariate linear equations:

$$\begin{aligned}\mathbf{Q}_0\eta+\sum_{m=1}^{7}c_m\mathbf{Q}_m\eta &= 0 \text{ for any } \eta\\ \text{or } \sum_{m=1}^{7}c_m\mathbf{Q}_m &= -\mathbf{Q}_0\end{aligned}$$

Write it in matrix forms,

$$\begin{bmatrix}\mathbf{Q}'_1 & \mathbf{Q}'_2 & \mathbf{Q}'_3 & \mathbf{Q}'_4 & \mathbf{Q}'_5 & \mathbf{Q}'_6 & \mathbf{Q}'_7\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\\c_4\\c_5\\c_6\\c_7\end{bmatrix}=-\mathbf{Q}'_0 \text{ or } \mathbf{Q}'\mathbf{c}=-\mathbf{Q}'_0$$

Plug in the numbers for this case, we have

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}_{6\times 7} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ -1 \\ 0 \end{bmatrix}_{6\times 1}$$

It can be easily shown that $\mathbf{Q}$ is not of full column rank. Therefore, the solutions for $\mathbf{c}$ is not unique, to solve this indefinite multivariate linear equations problem, we can implement fundamental transformations for the matrix $[\mathbf{Q}, -\mathbf{Q}_0']$.

$$trans\left([\mathbf{Q}, -\mathbf{Q}_0']\right) = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution for this system of equations is given by

$$\mathbf{c} = \begin{bmatrix} -1 + a_1 + a_3 \\ -1 + a_2 + a_3 \\ -a_1 - a_2 - a_3 \\ 1 - a_1 - a_2 - 2a_3 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Four bases for $\mathbf{c}$ are when $[a_1, a_2, a_3] = [0, 0, 0]$, $[1, 0, 0]$, $[0, 1, 0]$, and $[0, 0, 1]$. They are

$$\begin{aligned} \mathbf{c}_1 &= [-1, -1, 0, 1, 0, 0, 0]' \\ \mathbf{c}_2 &= [0, -1, -1, 0, 1, 0, 0]' \\ \mathbf{c}_3 &= [-1, 0, -1, 0, 0, 1, 0]' \\ \mathbf{c}_4 &= [0, 0, -1, -1, 0, 0, 1]' \end{aligned}$$

The corresponding transformations are

$$
\begin{aligned}
\mathbf{M}_{D,1a} &= \mathbf{I} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \mathbf{I}_T - \mathbf{I}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T + \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \bar{\mathbf{J}}_T \\
\mathbf{M}_{D,1b} &= \mathbf{I} - \mathbf{I}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \mathbf{I}_T + \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T \\
\mathbf{M}_{D,1c} &= \mathbf{I} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \mathbf{I}_T - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \mathbf{I}_T + \bar{\mathbf{J}}_{N_1} \otimes \bar{\mathbf{J}}_{N_2} \otimes \mathbf{I}_T \\
\mathbf{M}_{D,1d} &= \mathbf{I} - \mathbf{I}_{N_1} \otimes \bar{\mathbf{J}}_{N_2 T} - \bar{\mathbf{J}}_{N_1} \otimes \mathbf{I}_{N_2} \otimes \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_{N_1 N_2} \otimes \mathbf{I}_T + \mathbf{2}\bar{\mathbf{J}}_{N_1 N_2 T}
\end{aligned}
$$

which completes the proof. □

3.7.4 Proof of Proposition 3.3.2

Proof. Because of dropping the homoskedasticity assumption in Balaszi (2015), we cannot take $\mathbf{\Sigma}$ outside of the trace function. We have

$$
\mathbf{\Sigma\Gamma}' = \begin{bmatrix}
\begin{bmatrix}
0 & \sigma^2_{111} & \cdots & \rho^{T-3}\sigma^2_{111} & \rho^{T-2}\sigma^2_{111} \\
0 & 0 & \cdots & \rho^{T-4}\sigma^2_{112} & \rho^{T-3}\sigma^2_{112} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & \sigma^2_{11,T-1} \\
0 & 0 & \cdots & 0 & 0
\end{bmatrix} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \ddots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \begin{bmatrix}
0 & \sigma^2_{N_1N_21} & \cdots & \rho^{T-3}\sigma^2_{N_1N_21} & \rho^{T-2}\sigma^2_{N_1N_21} \\
0 & 0 & \cdots & \rho^{T-4}\sigma^2_{N_1N_22} & \rho^{T-3}\sigma^2_{N_1N_22} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & \sigma^2_{N_1N_2,T-1} \\
0 & 0 & \cdots & 0 & 0
\end{bmatrix}
\end{bmatrix}
$$

Now we have the same representaion for $\mathbf{\Sigma\Gamma}'$, the different expression form for $tr\left(\mathbf{\Sigma\Gamma}'\mathbf{M}_D\right)$ of different within estimators lies in $\mathbf{M}_D$. Before giving the proof, we denote $[\cdot]_{ij}$ the partition matrix with the partitioned row $[N_2T \times (i-1) + T \times (j-1) + 1 : N_2T \times (i-1) + T \times j]$ and the same partitioned column $[N_2T \times (i-1) + T \times (j-1) + 1 : N_2T \times (i-1) + T \times j]$.

Therefore, the $T \times T$-dimensional partition matrix $[\mathbf{\Sigma\Gamma}']_{ij}$ is

$$[\mathbf{\Sigma\Gamma}']_{ij} = \begin{bmatrix} 0 & \sigma^2_{ij1} & \cdots & \rho^{T-3}\sigma^2_{ij1} & \rho^{T-2}\sigma^2_{ij1} \\ 0 & 0 & \cdots & \rho^{T-4}\sigma^2_{ij2} & \rho^{T-3}\sigma^2_{ij2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma^2_{ij,T-1} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Firstly consider $\mathbf{M}_{D,1b} = \mathbf{I} - \mathbf{I}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T - \bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\mathbf{I}_T + \bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$, right multiplying $\mathbf{I}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$ means taking average along the each row for partitioned matrix $[\mathbf{\Sigma\Gamma}']_{ij}$, i.e., $[\mathbf{\Sigma\Gamma}']_{ij}\,\bar{\mathbf{J}}_T$.

$$[\mathbf{\Sigma\Gamma}']_{ij}\,\bar{\mathbf{J}}_T = \begin{bmatrix} \bar{\sigma}^2_{ij1} & \bar{\sigma}^2_{ij1} & \cdots & \bar{\sigma}^2_{ij1} & \bar{\sigma}^2_{ij1} \\ \bar{\sigma}^2_{ij2} & \bar{\sigma}^2_{ij2} & \cdots & \bar{\sigma}^2_{ij2} & \bar{\sigma}^2_{ij2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\sigma}^2_{ij,T-1} & \bar{\sigma}^2_{ij,T-1} & \cdots & \bar{\sigma}^2_{ij,T-1} & \bar{\sigma}^2_{ij,T-1} \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

where $\bar{\sigma}^2_{ijt} = \sigma^2_{ijt}\left(1 + \cdots + \rho^{T-t-1}\right)/T$. Right multiplying $\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\mathbf{I}_T$ has no effect on the diagonal items since taking averaging over i is equivalent to taking average over N_1 zero items for diagonal items. Right multiplying $\bar{\mathbf{J}}_{N_1}\otimes\mathbf{I}_{N_2}\otimes\bar{\mathbf{J}}_T$ means taking average along the each row for the stack matrix of partitioned matrix $[\mathbf{\Sigma\Gamma}']_{ij}$ and $N_1 - 1$ zero matrix with dimension $T \times T$, i.e., $\left[[\mathbf{\Sigma\Gamma}']_{ij}, \mathbf{0}_{T\times(N_1-1)T}\right]\mathbf{J}_{N_1T\times T}/\left(N_1T\right) = [\mathbf{\Sigma\Gamma}']_{ij}\,\bar{\mathbf{J}}_T/N_1$. Thus after transfomed by $\mathbf{M}_{D,1b}$, the partition matrix $[\mathbf{\Sigma\Gamma}']_{ij}$ becomes

$$[\mathbf{\Sigma\Gamma}'\mathbf{M}_{D,1b}]_{ij} = \begin{bmatrix} -\bar{\sigma}^2_{ij1} + \bar{\sigma}^2_{ij1}/N_1 & \sigma^2_{ij1} - \bar{\sigma}^2_{ij1} + \bar{\sigma}^2_{ij1}/N_1 & \cdots & \rho^{T-2}\sigma^2_{ij1} - \bar{\sigma}^2_{ij1} + \bar{\sigma}^2_{ij1}/N_1 \\ 0 & -\bar{\sigma}^2_{ij2} + \bar{\sigma}^2_{ij2}/N_1 & \cdots & \rho^{T-3}\sigma^2_{ij2} - \bar{\sigma}^2_{ij2} + \bar{\sigma}^2_{ij2}/N_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{ij,T-1} - \bar{\sigma}^2_{ij,T-1} + \bar{\sigma}^2_{ij,T-1}/N_1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Therefore,

$$tr\left(\mathbf{\Sigma\Gamma}'\mathbf{M}_{D,1b}\right) = \left[-\frac{1}{T\left(1-\rho\right)}\bar{\sigma} + \frac{1}{T^2\left(1-\rho\right)}\bar{\sigma}_\rho\right]\left(1 - \frac{1}{N_1}\right)$$

which completes the proof of $\mathbf{Bias}_{1b}$. After trivial calculation, we can also obtain the asymptotic distribution of $\hat{\theta}_{1b}$. The proofs of other biases and the proofs of the asymptotic distributions of other estimators closely follow this approach. □

3.7.5 Proof of Theorem 3.4.1

Proof. This proof closely follows the proof of Dhaene and Jochmans (2015). In their framework, they let N and T go to infinity in the same rate, which is equivalent to letting $N_1 \times N_2$ and T go to infinity in the same rate. We can allow N_1, N_2, and T go to infinity at the same rate because we consider a situation in which our condition is stronger than Assumption 3.3 of DJ15. Note that

$$\tilde{\theta}_{1/2} = \frac{1}{2}\left(\tilde{\theta}_{\mathcal{S}_1} + \tilde{\theta}_{\mathcal{S}_2}\right), \qquad \tilde{\theta}_{\mathcal{S}_m} \equiv 2\hat{\theta} - \bar{\theta}_{\mathcal{S}_m}$$

where $\bar{\theta}_{\mathcal{S}_m} = \sum_{S_{mk}\in\mathcal{S}_m} \frac{|S_{mk}|}{T}\hat{\theta}_{S_{mk}}$. In particular, $\bar{\theta}_{\mathcal{S}_1} = \frac{\lceil T/2 \rceil}{T}\hat{\theta}_{S_{11}} + \frac{T-\lceil T/2 \rceil}{T}\hat{\theta}_{S_{12}}$ and $\bar{\theta}_{\mathcal{S}_2} = \frac{\lfloor T/2 \rfloor}{T}\hat{\theta}_{S_{21}} + \frac{T-\lfloor T/2 \rfloor}{T}\hat{\theta}_{S_{22}}$. Averaging over the equivalence of $\mathcal{S}$ does not affect the asymptotic properties and $\tilde{\theta}$. Thus it suffices to consider the asymptotic behavior of $\tilde{\theta}_{\mathcal{S}}$. Following the proof of Theorem 3.1 in DJ15, we have the result

$$\sqrt{N_1N_2T}\left(\tilde{\theta}_{1/2} - p\lim_{N\to\infty}\tilde{\theta}_{1/2}\right) \to_d N(0, \Sigma^{-1})$$

For all the fixed effect estimators in our asymptotic pattern, we have that $\hat{\theta} - \theta = \frac{\tilde{\mathbf{B}}_1}{T} + \frac{\tilde{\mathbf{B}}_2}{T^2} + o_p\left(\frac{1}{T^2}\right)$ where $\tilde{\mathbf{B}}_1$ and $\tilde{\mathbf{B}}_2$ are $O_p(1)$ and they could be zero. For any $\hat{\theta}_{S_{mk}}$, we also have $\hat{\theta}_{S_{mk}} - \theta = \frac{\tilde{\mathbf{B}}_1}{T_{mk}} + \frac{\tilde{\mathbf{B}}_2}{T_{mk}^2} + o_p\left(\frac{1}{T_{mk}^2}\right)$. Then

$$\bar{\theta}_{\mathcal{S}_m} - \theta = \frac{2\tilde{\mathbf{B}}_1}{T} + \frac{\tilde{\mathbf{B}}_2}{TT_{mk}} + \frac{\tilde{\mathbf{B}}_2}{T\left(T - T_{mk}\right)} + o_p\left(\frac{1}{T^2}\right)$$

By the construction of $\tilde{\theta}_{1/2}$

$$\tilde{\theta}_{1/2} - \theta = \frac{2\tilde{\mathbf{B}}_2}{T^2} - \frac{\tilde{\mathbf{B}}_2}{TT_{mk}} - \frac{\tilde{\mathbf{B}}_2}{T\left(T - T_{mk}\right)} = O\left(T^{-2}\right)$$

and Propositions in Section 3, we have

$$\sqrt{N_1N_2T}(p\lim_{N\to\infty}\tilde{\theta}_{1/2} - \theta) = \sqrt{N_1N_2T}O(T^{-2}) \to 0$$

provided $N_1, N_2, T \to \infty$ with $N_1/T \to \delta_1$ and $N_2/T \to \delta_2$. Therefore, the bias is asymptotically negligible. Now we can directly get the asymptotic distribution

$$\sqrt{N_1 N_2 T}\left(\tilde{\theta}_{1/2} - \theta\right) \to_d N\left(\mathbf{0}, \mathbf{\Phi}^{-1}\mathbf{\Omega}\mathbf{\Phi}^{-1}\right)$$

This completes the proof. □

Bibliography

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics 7*(1), 22–53.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1741–1768.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics 10*(4), 1747–1785.

Byrne, D. P., A. L. Nauze, and L. A. Martin (2018). Tell me something I don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics 100*(3), 510–527.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.

Chong, A., I. Cohen, E. Field, E. Nakasone, and M. Torero (2016). Iron deficiency and schooling attainment in Peru. *American Economic Journal: Applied Economics 8*(4), 222–55.

Crépon, B., F. Devoto, E. Duflo, and W. Parienté (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics 7*(1), 123–50.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics 2*(2), 267–277.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics 4*, 3895–3962.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics 128*(4), 1499–1545.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika 58*(3), 403–417.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*(1), 259–276.

Hahn, J., K. Hirano, and D. Karlan (2011). Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics 29*(1), 96–108.

Hu, Y. (2016). Generalized Efron's biased coin design and its theoretical properties. *Journal of Applied Probability 53*(2), 327–340.

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics 40*(3), 1794–1815.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis 100*(8), 1816–1829.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *The Annals of Statistics 26*(2), 755–770.

Ma, W., Y. Qin, Y. Li, and F. Hu (2018). Statistical inference of covariate-adjusted randomized experiments. *arXiv preprint arXiv:1807.09678*.

Montgomery-Smith, S. J. (1993). Comparison of sums of independent identically distributed random variables. *Probability and Mathematical Statistics 14*(2), 281–285.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy 119*(1), 39–77.

Peña, V. H., T. L. Lai, and Q.-M. Shao (2008). *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science & Business Media.

Shao, J. and X. Yu (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics 69*(4), 960–969.

Shao, J., X. Yu, and B. Zhong (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*(2), 347–360.

Tabord-Meehan, M. (2018). Stratification trees for adaptive randomization in randomized controlled trials. *arXiv preprint arXiv:1806.05127*.

van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wei, L. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association 73*(363), 559–563.

Zhang, Y. (2018). Extremal quantile treatment effects. *The Annals of Statistics 46*(6B), 3707–3740.

Bibliography for Chapter 2

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.

Anderson, J.E., and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93, 171-192.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20, 105-134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443-1458.

Bosquet, C. and Boulhol, H. (2009). Gravity, log of gravity and the “distance puzzle”. *Working Paper*, Université Aix-Marseille.

Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65, 1129-1151.

De Jong, R. M. (1996). The Bierens test under data dependence. *Journal of Econometrics*, 72, 1-32.

Delgado, M.A. and Manteiga, W.G. (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics*, 29, 1469-1507.

Dunford, N. and Schwartz, J. T. (1963). *Linear Operators, Part 2: Spectral theory, Self adjoint operators in Hilbert space.* John Wiley & Sons Inc., New York.

Eubank, R. and Hart, J. (1992). Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics*, 20, 1412–1425.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, 153-193.

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory*, 14, 604-621.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, 64, 865-890.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests. *Journal of Nonparametric Statistics*, 10, 245-274.

Fan, Y. and Li, Q. (2000). Consistent model specification tests. *Econometric Theory*, 16, 1016-1041.

Feenstra, R.C., Markusen, J.R., and Rose, A.K. (2001). Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics*, 19, 231-247.

Glick, R., and Rose, A. K. (2002). Does a currency union affect trade? The time-series evidence. *European Economic Review*, 46, 1125-1151.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Annals of Statistics*, 7, 110-123.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 1926-1947.

Hausman, J. A. (1978). Specification testing in econometrics. *Econometrica*, 46, 1251-1271.

Henderson, D. J., and Millimet, D. L. (2008). Is gravity linear? *Journal of Applied Econometrics*, 23, 137-172.

Hong, Y. and White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica*, 63, 1133-1159.

Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69, 599-631.

Hsiao, C., Li, Q., and Racine, J. S. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics*, 140, 802-826.

Huang, M., Sun, Y., and White, H. (2016). A flexible nonparametric test for conditional independence. *Econometric Theory*, 32, 1434-1482.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics*, 40, 633-643.

Kepaptsoglou, K., Karlaftis, M. G., and Tsamboulas, D. (2010). The gravity model specification for modeling international trade flows and free trade agreement effects: a 10-year review of empirical studies. *The Open Economics Journal*, 3, 1-13.

Lee, J. (1990). *U-statistics: Theory and Practice*. CRC Press, New York.

Li, Q., Hsiao, C., and Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics*, 112, 295-325.

Li, Q. and Racine, J.S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.

Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, 87, 145-165.

Li, R., Zhong, W., and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107, 1129-1139.

Linton, O., Gozalo, P. (2014). Testing conditional independence restrictions. *Econometric Reviews*, 33, 523-552.

Millimet, D. L., List, J. A., and Stengos, T. (2003). The environmental Kuznets curve: Real progress or misspecified models? *Review of Economics and Statistics*, 85, 1038-1047.

Millimet, D. L., and Osang, T. (2007). Do state borders matter for US intranational trade? The role of history and internal migration. *Canadian Journal of Economics*, 40, 93-126.

Martinez-Zarzoso, I. (2013). The log of gravity revisited. *Applied Economics*, 45, 311-327.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics* 29: 229-256.

Papke, L.E., and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11, 619-632.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society, Series B*, 31, 350-371.

Robinson, P. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies*, 56, 511-534.

Rose, A. K. (2005). Does the WTO make trade more stable?. *Open Economies Review*, 16, 7-22.

Santos Silva, J.M.C., and Tenreyro, S. (2006). The log of gravity. *Review of Economics and statistics,* 88, 641-658.

Shao, X., and Zhang, J. (2014). Martingale divergence correlation and its use in high dimensional variable screening. *Journal of the American Statistical Association*, 109, 1302-1318.

Song, K. (2009). Testing conditional independence via Rosenblatt transform. *Annals of Statistics*, 37, 4011-4045.

Stinchcombe, M.,White, H. (1998). Consistent specification testing with nuisance parameters present only under alternative. *Econometric Theory* 14: 295-324.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25, 613-641.

Su, L., Hoderlein, S., and H. White (2015). Testing monotonicity in unobservables with panel data. *Working Paper*, Singapore Management University.

Su, L., Jin, S., and Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*, 186, 222-244.

Su, L., and Lu, X. (2013). Nonparametric dynamic panel data models: kernel estimation and specification testing. *Journal of Econometrics*, 176, 112-133.

Su, L., and Qu, X. (2017). Specification test for spatial autoregressive models. *Journal of Business & Economic Statistics*, 35(4), 572-584.

Su, L., and White, H. (2007). A consistent characteristic function-based test for conditional independence. *Journal of Econometrics*, 141, 807-834.

Su, L., and White, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory,* 24, 829-864.

Su, L., and White, H. (2014). Testing conditional independence via empirical likelihood. *Journal of Econometrics* 182, 27-44..

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35, 2769-2794.

Székely, G. J., and Rizzo, M. L. (2009). Brownian distance covariance. *Annals of Applied Statistics*, 3, 1236-1265.

Székely, G. J., and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Annals of Statistics*, 42, 2382-2412.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy.* The Twentieth Century Fund, New York.

Whang, Y.J. (2000). Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics*, 98, 27-46.

Wooldridge, J. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory*, 8, 452–475.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9, 501-513.

Wu, C.-F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261-1295.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory*, 8, 435-451.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75, 263-289.

Bibliography for Chapter 3

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), pp.277-297.

Bai, J. (2013). Fixed-Effects Dynamic Panel Models, a Factor Analytical Method. *Econometrica*, 81(1), 285-314.

Balazsi, L., Matyas, L., and Wansbeek, T. (2018). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, 37(3), 212-227.

Claeskens, G., and Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98(464), 900-916.

Dhaene, G. and Jochmans, K., 2015. Split-Panel Jackknife Estimation of Fixed-Effect Models. *The Review of Economic Studies*, 82(3), pp.991-1030.

Hausman, J.A. and Taylor, W.E., 1981. Panel Data and Unobservable Individual Effects. *Econometrica*, pp.1377-1398.

Hsiao, C., 2014. Analysis of Panel Data (No. 54). Cambridge university press.

Lee, Y., and Phillips, P. C. (2015). Model selection in the presence of incidental parameters. *Journal of Econometrics*, 188(2), 474-489.

Lu, X., K. Miao, and L. Su. (2020). Specification of Fixed Effects in Three-Dimensional Panels. *Econometrics Reviews*, forthcoming.

Matyas, L. (1997). Proper Econometric Specification of The Gravity Model. *The World Economy*, 20(3), pp.363-368.

Matyas, L. (2017). The econometrics of multi-dimensional panels. Berlin: Springer, doi, 10, 978-3.

Moon, H. R., and Weidner, M. (2017). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *Econometric Theory*, 33(1), 158-195.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica: Journal of the Econometric Society*, pp.1417-1426.

Phillips, P.C. and Moon, H.R., 1999. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, 67(5), pp.1057-1111.

White, H. (1984). Asymptotic Theory for Econometricians. Academic press.