



Theory article

Model averaging with causal effects for partially linear models

Xiaowei Zhang^{1,*} and Junliang Li²

¹ Department of Statistics and Data Science, School of Economics, Jinan University, Guangzhou, 510632, China

² The second middle school of yuncheng county, Heze city of Shandong Province, 274700, China

* **Correspondence:** Email: xwzhang@stu2021.jnu.edu.cn.

Abstract: Treatment effects with heterogeneity and heteroskedasticity are widely studied and applied in many fields, such as statistics and econometrics. The conditional average treatment effect provides an excellent measure of the heterogeneous treatment effect. In this paper, we propose a model averaging estimation for the conditional average treatment effect with partially linear models based on the jackknife-type criterion under heteroscedastic error. Within this context, we provide theoretical justification for our model averaging approach, and we establish asymptotic optimality and weight convergence properties for our model under certain conditions. The performance of our proposed estimator is compared with that of classical estimators by using a Monte Carlo study and empirical analysis.

Keywords: model averaging; causal inference; partially linear models; conditional average treatment effect; Jackknife-type criterion

Mathematics Subject Classification: 62F40, 62G20, 62G99

1. Introduction

Causal effects are fundamental to project evaluation across various fields, such as economics, finance, and biomedicine. It is well known that the central issue in project evaluation in these fields is to identify the “causal relationships” that exist between project treatments and project outcomes and to quantify the “causal effects”. For example, pharmaceutical companies are interested in the effect of a new drug or device developed to treat a disease, and investment banks are concerned with the profitability of companies that have received significant capital investment. These research questions rely on estimating treatment effects; thus, studying the identification, estimation, and empirical application of treatment effects is crucial and meaningful. Causal effects are also referred to as treatment effects in the literature.

It is essential, in some cases, to accurately capture heterogeneous treatment effects. For example, a new medicine could be more beneficial for children than for adults; an advertising strategy for sanitary napkins may be more persuasive for women than for men. These results indicate that using knowledge of heterogeneous treatment effects can maximize the value and effectiveness of treatment programs [15]. A commonly used heterogeneous treatment effects measurement is the conditional average treatment effect (CATE), to which much recent attention has been given.

In several studies, CATE has been estimated by fitting a parametric model to the relationship between observations and baseline covariates, in conjunction with treatment assignment. Alternatively, CATE can be estimated by modeling the conditional mean of potential outcomes using a parametric model based on the baseline covariates for each treatment group. Examples include [2, 9, 19, 22]. However, in practice, practitioners generally have datasets in which not all covariates show a linear relationship with each other. Instead, the data may be nonparametric. In this context, partially linear models (PLMs) are more applicable to such datasets, which combine the interpretability of linear models with the flexibility of nonparametric models.

Therefore, we consider a partially linear regression framework [3] in which the distribution of the response Y_i may depend on a binary treatment assignment indicator $\delta_i \in \{0, 1\}$ and the baseline covariates (X_i, U_i) . To segregate the treatment differences of the primary interest, we assume for the observations that

$$Y_{t,i} - Y_{c,i} = \mu_i + e_i = \sum_{j=1}^{\infty} x_{ij}\beta_j + g(U_i) + e_i, \quad i = 1, \dots, n, \quad (1.1)$$

where $\{Y_{t,i}, Y_{c,i}\}$ denote the Y_i associated pair of potential outcomes and δ_i represents the treatment indicator variable, taking $\delta_i = 1$ if the individual belongs to the treatment group and $\delta_i = 0$ otherwise. $X_i = (x_{i1}, x_{i2}, \dots)^T$ is covariate and U_i is univariate covariate. β is an unknown coefficient vector associated with X_i , $g(\cdot)$ is an unknown smooth nonlinear function, and e_1, \dots, e_n are unobservable heteroscedastic random errors independent of $\{X_i, U_i\}_{i=1}^n$ with conditional mean $E(e_i|X_i, U_i) = 0$ and conditional variance $E(e_i^2|X_i, U_i) = \sigma_i^2$. A fundamental problem in statistics and causal inference is to improve predictive accuracy based on observable datasets $\{Y_i, X_i, U_i, \delta_i\}_{i=1}^n$.

Researchers, however, usually have multiple candidate models assumed based on various understandings of the observed data. Thus, Polling and Yang [15] proposed the treatment effect across-validation (TECV) method. It is a model selection method for studying CATE, which is able to identify a model that is most suitable for estimating CATE among multiple candidate models. However, there is a risk of model uncertainty, and it is impossible to know whether the model is the most suitable for the dataset. If researchers cannot select the optimal model, they will face missing useful information, leading to unstable estimation. The model averaging method can reduce the risk of regression estimation and the bias introduced by selecting a single model, avoid ignoring the useful information contained in the remaining candidate models, and thereby improve prediction accuracy.

Model averaging methods have been applied to handle causal inference problems. For example, Gao et al., [5] developed a model averaging method based on the JMA of [7] to estimate the average treatment effect (ATE). [11] proposed a data-driven approach to averaging the estimators over candidate specifications to address the specification uncertainty in the weighted estimation of propensity scores for the average treatment effect on treated (ATT). Rolling et al., [16] introduced a model combination technique, treatment effect estimation by mixing (TEEM), which was designed to amalgamate estimators from various programs. This approach is used to generate more accurate

CATE estimations. Although TEEM is compatible with parametric, nonparametric, or semiparametric statistical models, as well as nonstatistical machine learning processes and even subjective expert judgment, the approach proposed by [16] may encounter difficulties in finding treatment-control pairs in each cell after partitioning. This is especially true when dealing with a large or even moderate number of covariates. Thus, we discussed a model averaging estimation for the CATE with multiple candidate partially linear regression models.

To the best of our knowledge, no optimal model averaging estimation has been developed for PLMs based on the jackknife-type criterion to address causal inference problems. Motivated by this, under heteroscedastic error, the primary goal of the current article is to develop a model averaging estimation for the conditional average treatment effect with PLMs based on a jackknife-type criterion called the CPLJMA method, where C, PL, and JMA represent causal effects, partially linear, and jackknife model averaging, respectively. However, it is difficult to directly extend the available results to our idea for the following reasons: For existing optimal model averaging estimations for PLMs, important examples include [27], who proposed optimal model averaging estimation for PLMs with the Mallows-type criterion based on kernel estimation (MAPLM). Zeng et al., [25] proposed focused information criteria and frequentist model averaging estimators for semiparametric partially linear models with missing responses and established their theoretical properties. We utilize a different jackknife-type weight choice criterion from theirs, and, additionally, we add an extra theorem of weight convergence. Our main contributions to these challenges are as follows: (i) On a theoretical basis, the proposed estimator is asymptotically optimal in terms of minimizing the squared error loss. (ii) The convergence property of the weights is investigated, and we prove that the sum of the weights assigned to the correct candidate model is one as the sample size increases to infinity. This is true, provided that there is at least one correctly specified candidate model. In the simulation section, numerical simulations and empirical analysis are applied to verify the validity of the proposed model averaging approach and the analytical framework; this proves theoretical support for the wide application of such model averaging methods.

The remainder of this paper is organized as follows. In Section 2, we describe the estimation procedure of the jackknife criterion for CATE based on PLMs, and we study its theoretical properties in Section 3. Section 4 illustrates the performance of our proposal via simulations and data examples. Concluding remarks are made in Section 5. Technical proofs are deferred to the Supplementary.

2. Model and estimation procedure

2.1. Preliminaries

First, we use the potential outcomes framework [8, 18] to define the CATE as the expectation of the random variable conditional on the observed value of the baseline covariates (X_i, U_i) , that is

$$\mu_i = E(Y_{t,i} - Y_{c,i} | X_i, U_i),$$

where $Y_{t,i} - Y_{c,i}$ indicates the treatment effect for the i th individual; this is also labeled the “fundamental problem of causal inference” because $Y_{t,i}$ and $Y_{c,i}$ are infeasible to observe simultaneously. The main goal of this study is to estimate μ_i based on the model averaging method.

Under the causal inference framework, we make the following identifiability assumptions [1]:

Assumption 1. *Consistency:* $Y_i = \delta_i Y_{t,i} + (1 - \delta_i) Y_{c,i}$;

Assumption 2. Unconfoundedness: $\{Y_{t,i}, Y_{c,i}\} \perp \delta_i \mid \{X_i, U_i\}$;

Assumption 3. Positivity: $0 < c_\pi \leq \pi(X_i, U_i) \leq 1 - c_\pi < 1$ almost surely, where $\pi(X_i, U_i) = P(\delta_i = 1 \mid Y_{t,i}, Y_{c,i}, X_i, U_i) = P(\delta_i = 1 \mid X_i, U_i)$ denotes the propensity score and c_π is a positive constant.

Assumption 1 links potential outcomes to observed outcomes and requires the potential outcomes to be well defined. Assumption 2 is conditional independence, which implies that the treatment assignment indicator δ_i is independent of the potential outcome $\{Y_{t,i}, Y_{c,i}\}$ given covariates X_i and U_i , and it requires that all potential confounding information on the relationship between treatments and potential outcomes is observed in the covariates; it precludes potential confounding between treatment assignments and outcomes. Assumption 3 implies that treatment assignments are not deterministic, which is crucial for controlling confounding bias and systematic differences between the treatment and control groups. This assumption also guarantees $\pi(X_i, U_i)$ and $1 - \pi(X_i, U_i)$ are invertible with probability one. [17] referred to the combination of Assumptions 2 and 3 as a “strongly ignorable treatment assignment”.

Inverse probability weighting (IPW) based on the potential outcome framework is a powerful tool for correcting confounding bias. The IPW approach utilizes the inverse of the propensity scores to construct weights for observed outcomes to balance baseline covariates between groups [10]. Under the strongly ignorable treatment assignment assumption, suppose

$$Z_{\pi,i} = \frac{\delta_i Y_i}{\pi(X_i, U_i)} - \frac{(1 - \delta_i) Y_i}{1 - \pi(X_i, U_i)}$$

to construct the unbiased estimator of μ_i given the conditions X_i, U_i . Then, after careful calculation, we have

$$\begin{aligned} E(Z_{\pi,i} \mid X_i, U_i) &= E \left[\frac{\delta_i Y_i}{\pi(X_i, U_i)} - \frac{(1 - \delta_i) Y_i}{1 - \pi(X_i, U_i)} \mid X_i, U_i \right] \\ &= \frac{E(\delta_i \mid X_i, U_i)}{\pi(X_i, U_i)} E(Y_{t,i} \mid X_i, U_i) - \frac{E(1 - \delta_i \mid X_i, U_i)}{1 - \pi(X_i, U_i)} E(Y_{c,i} \mid X_i, U_i) \\ &= E(Y_{t,i} \mid X_i, U_i) - E(Y_{c,i} \mid X_i, U_i) \\ &= \mu_i. \end{aligned} \tag{2.1}$$

The unreasonable model selected to estimate the CATE may yield uncertain estimations. To provide a more reasonable and robust CATE estimator, we develop a model averaging estimation for CATE with multiple candidate PLMs.

2.2. Estimation procedure

As discussed earlier, we can obtain computationally feasible PLMs of heterogeneous causal effects based on treatment-control pairs. In accordance with Eq (2.1) and model (1.1), we let $e_{\pi,i} = Z_{\pi,i} - \mu_i$. Then, we have

$$Z_{\pi,i} = \mu_i + e_{\pi,i} = \sum_{j=1}^{\infty} x_{ij} \beta_j + g(U_i) + e_{\pi,i}, \tag{2.2}$$

where μ_i denotes the CATE for the i th individual, $E(e_{\pi,i} \mid X_i, U_i) = 0$, and $E(e_{\pi,i}^2 \mid X_i, U_i) = \sigma_{\pi,i}^2$. Let us note that $\{Z_{\pi,i}, X_i, U_i\}_{i=1}^n$ is fully observed when $\pi(X_i, U_i)$ is known.

Specifically, we consider multiple candidate models for model (2.2) of the form

$$\mu_i^{(m)} = \sum_{j=1}^{k_m} x_{ij}^{(m)} \beta_j^{(m)} + g(U_i), \quad m = 1, \dots, M_n, \quad (2.3)$$

used for evaluating μ_i , where $x_{ij}^{(m)}$ is the j th entry of $X_i^{(m)}$, $X_i^{(m)}$ is an k_m -dimensional subvector of X_i , $\beta_j^{(m)}$ is the corresponding regression coefficient vector, $g(\cdot)$ is an unknown function in the nonparametric part, and M_n denotes the total number of candidate models allowed to go to infinity.

Define $\mathbf{Z}_\pi = (Z_{\pi,1}, \dots, Z_{\pi,n})^T \in \mathbb{R}^n$, $\mathbf{X}^{(m)} = (X_1^{(m)}, \dots, X_n^{(m)})^T \in \mathbb{R}^{n \times k_m}$, where $\{x_{ij}^{(m)}\}_{i=1}^n$ is an $1 \times k_m$ column vector, $\mathbf{g}(\mathbf{U}) = (g(U_1), \dots, g(U_n))^T \in \mathbb{R}^n$, and $\mathbf{e}_\pi = (e_{\pi,1}, \dots, e_{\pi,n})^T \in \mathbb{R}^n$. Then, the m th candidate model in matrix form is

$$\mathbf{Z}_\pi = \mathbf{X}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{g}(\mathbf{U}) + \mathbf{e}_\pi.$$

To estimate the nonparametric function, we use the B-spline regression method. Let \mathcal{S}_n be the space of polynomial splines of degree $l \geq 1$, and let $\{\psi_k, k = 1, \dots, d_n\}$ denote a normalized B-spline basis. For any $g_n \in \mathcal{S}_n$, we have

$$\mathbf{g}_n(\mathbf{U}) = \sum_{k=1}^{d_n} \psi_k(\mathbf{U}) \alpha_k = \Psi^T(\mathbf{U}) \boldsymbol{\alpha},$$

for some coefficients $\{\alpha_k\}_{k=1}^{d_n}$, where $\Psi(\mathbf{U}) = (\psi_1(\mathbf{U}), \dots, \psi_{d_n}(\mathbf{U}))^T$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{d_n})^T$. Here, d_n increases with n . We define the $n \times d_n$ matrix $\mathbf{K} = (\Psi(U_1), \dots, \Psi(U_n))^T$. Then, we assume that the $n \times (p + d_n)$ matrix $\mathbf{X}^{(m)*} = (\mathbf{X}^{(m)}, \mathbf{K})$ is nonsingular and associated with the unknown $(p + d_n)$ -dimensional parameter vector $\boldsymbol{\gamma}^{(m)} = (\boldsymbol{\beta}^{(m)*}, \boldsymbol{\alpha}^T)^T$. Thus, we have

$$\mu_{\pi,n}^{(m)} = \mathbf{X}^{(m)*} \boldsymbol{\gamma}^{(m)} = \mathbf{X}^{(m)} \boldsymbol{\beta}^{(m)} + \mathbf{K} \boldsymbol{\alpha}.$$

By regressing \mathbf{Z}_π on $\mathbf{X}^{(m)*}$, the least squares estimator of $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\alpha}$ can be obtained as

$$\hat{\boldsymbol{\beta}}^{(m)} = \{\mathbf{X}^{(m)*T} (\mathbf{I} - \mathbf{Q}) \mathbf{X}^{(m)*}\}^{-1} \mathbf{X}^{(m)*T} (\mathbf{I} - \mathbf{Q}) \mathbf{Z}_\pi,$$

and

$$\hat{\boldsymbol{\alpha}} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T (\mathbf{Z}_\pi - \mathbf{X}^{(m)} \hat{\boldsymbol{\beta}}^{(m)}),$$

where $\mathbf{Q} = \mathbf{K}(\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T$ is a symmetric idempotent matrix. Then

$$\hat{\mu}_\pi^{(m)} = \mathbf{X}^{(m)} \hat{\boldsymbol{\beta}}^{(m)} + \mathbf{K} \hat{\boldsymbol{\alpha}} = \{\mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)*T} \tilde{\mathbf{X}}^{(m)*})^{-1} \tilde{\mathbf{X}}^{(m)*T}\} \mathbf{Z}_\pi = \mathbf{P}^{(m)} \mathbf{Z}_\pi,$$

where $\tilde{\mathbf{X}}^{(m)} = (\mathbf{I} - \mathbf{Q}) \mathbf{X}^{(m)}$ is a symmetric idempotent matrix and $\mathbf{P}^{(m)} = \mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)*T} \tilde{\mathbf{X}}^{(m)*})^{-1} \tilde{\mathbf{X}}^{(m)*T}$. Then the corresponding model averaging estimator of $\boldsymbol{\mu}$ can be formulated as

$$\hat{\boldsymbol{\mu}}_\pi(\boldsymbol{\omega}) = \sum_{m=1}^{M_n} \omega_m \hat{\mu}_\pi^{(m)} = \mathbf{P}(\boldsymbol{\omega}) \mathbf{Z}_\pi, \quad (2.4)$$

where $\mathbf{P}(\boldsymbol{\omega}) = \sum_{m=1}^{M_n} \omega_m \mathbf{P}^{(m)}$ and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{M_n})^T$ is a weight vector belonging to the continuous set $\mathcal{H}_n = \{\boldsymbol{\omega} \in [0, 1]^{M_n} : \sum_{m=1}^{M_n} \omega_m = 1\}$.

Notably, the choice of weight vector is crucial in the model averaging method. Thus, we consider the jackknife-type criterion to choose the weight vector ω for (2.4) in the PLMs framework. Specifically, leave-one-out cross-validation (LOO-CV) is used to estimate μ , and then the estimator in the m th candidate model is given by

$$\tilde{\mu}_\pi^{(m)} = \tilde{\mathbf{P}}^{(m)} \mathbf{Z}_\pi \quad \text{and} \quad \tilde{\mathbf{P}}^{(m)} = \mathbf{P}^{(m)} - \mathbf{D}^{(m)} \mathbf{A}^{(m)},$$

where $\mathbf{D}^{(m)} = \text{diag}(D_{11}^{(m)}, \dots, D_{nn}^{(m)}) \in \mathbb{R}^{n \times n}$ with the i th diagonal element $D_{ii}^{(m)} = h_{m,ii}/(1 - h_{m,ii})$, $\mathbf{A}^{(m)} = \mathbf{I}_n - \mathbf{P}^{(m)}$, and $h_{m,ii}$ is the i th diagonal entry of $\mathbf{P}^{(m)}$. Thus, the jackknife-type model averaging estimator is

$$\tilde{\mu}_\pi(\omega) = \sum_{m=1}^{M_n} \omega_m \tilde{\mu}_\pi^{(m)} = \tilde{\mathbf{P}}(\omega) \mathbf{Z}_\pi,$$

where $\tilde{\mathbf{P}} = \sum_{m=1}^{M_n} \omega_m \tilde{\mathbf{P}}^{(m)}$. Then, the weight choice criterion is

$$CV_\pi(\omega) = \|\mathbf{Z}_\pi - \tilde{\mu}_\pi(\omega)\|^2. \quad (2.5)$$

The optimal weight vector is obtained by minimizing the criterion in (2.5) over the space \mathcal{H}_n . However, such a minimization process in real-world data analysis is computationally infeasible because $\pi(X_i, U_i)$ is generally unknown. In our modeling framework, we estimate it by adopting the logistic partially linear models (LPLMs) in [24],

$$\hat{\pi}(X_i, U_i) = \frac{e^{X_i^T \theta + \kappa(U_i)}}{1 + e^{X_i^T \theta + \kappa(U_i)}}, \quad i = 1, \dots, n \quad (2.6)$$

which relies on generalized partially linear models (GPLMs) in [14], where the coefficients θ for the linear part and the nonparametric part $\kappa(\cdot)$ are estimated using B-spline basis estimation. This method is facilitated by the “sgplm1” function within the **R** package “gplm”, with the degrees of freedom set to “ $df = 3$ ”. Further elucidation of this choice is provided in the subsequent numerical simulations. Then, $\hat{\pi}(X_i, U_i)$ is substituted for $\pi(X_i, U_i)$ with \mathbf{Z}_π to obtain $\mathbf{Z}_{\hat{\pi}}$. Then, a feasible counterpart of $CV_\pi(\omega)$ in (2.5) becomes

$$CV_{\hat{\pi}}(\omega) = \|\mathbf{Z}_{\hat{\pi}} - \tilde{\mu}_{\hat{\pi}}(\omega)\|^2.$$

The optimal weights $\hat{\omega}_{cv}$ are obtained by selecting $\omega \in \mathcal{H}_n$ to minimize the jackknife-type criterion in

$$\hat{\omega}_{cv} = \arg \min_{\omega \in \mathcal{H}_n} CV_{\hat{\pi}}(\omega).$$

Given $\tilde{\omega}_{cv}$, substituting it into (2.4) yields the optimal model averaging estimator of μ as $\hat{\mu}_{\hat{\pi}}(\hat{\omega}_{cv})$ in the case where $\hat{\pi}(X_i, U_i)$. $CV_{\hat{\pi}}(\omega)$ is a quadratic programming problem with respect to ω , that is, $CV_{\hat{\pi}}(\omega) = \omega^T \mathbf{H}_{\hat{\pi}}^T \mathbf{H}_{\hat{\pi}} \omega$, where $\mathbf{H}_{\hat{\pi}} = (\mathbf{Z}_{\hat{\pi}} - \tilde{\mathbf{P}}_{(1)} \mathbf{Z}_{\hat{\pi}}, \dots, \mathbf{Z}_{\hat{\pi}} - \tilde{\mathbf{P}}_{(M_n)} \mathbf{Z}_{\hat{\pi}})$.

3. Theoretical properties

In this section, we focus on some theoretical properties of our proposed model averaging method. In Subsection 3.1, we prove the asymptotic optimality of the model averaging estimator $\hat{\mu}(\hat{\omega}_{cv})$ by illustrating that the selected weight vector $\hat{\omega}_{cv}$ yields a squared error that is asymptotically identical to that of the infeasible optimal weight vector. Subsection 3.2 concerns the convergence property of the optimal weight vector $\hat{\omega}_{cv}$. When the sample size tends to infinity, the sum of the weights assigned to the correct model by the optimal weight vector obtained by the proposed method converges to one in probability.

3.1. Asymptotic optimality

Before introducing the theoretical properties, we first define some notations. Let $\mu_{t,i} = E(Y_{t,i}|X_i, U_i)$ and $e_{t,i} = Y_{t,i} - \mu_{t,i}$ denote the conditional expectation and the random error of the treatment group, respectively, in which $\sigma_{t,i}^2 = E(e_{t,i}^2|X_i, U_i)$. The loss function and the corresponding risk function of $\hat{\mu}_\pi(\omega)$ are defined as

$$L_\pi(\omega) = \|\mu - \hat{\mu}_\pi(\omega)\|^2 \quad \text{and} \quad R_\pi(\omega) = E\{L_\pi(\omega)|X_i, U_i\},$$

respectively, where $\|\cdot\|$ is the Euclidean norm. Let $\xi_\pi = \inf_{\omega \in \mathcal{H}_n} R_\pi(\omega)$, $\bar{k} = \max_{1 \leq m \leq M_n} k_m$, and $\bar{h} = \max_{1 \leq m \leq M_n} \max_{1 \leq i \leq n} h_{m,ii}$. The following conditions are assumed with respect to $n \rightarrow \infty$.

(C1) $\sqrt{n}\|\hat{\theta}_n - \theta_0\| = O_p(1)$ and $\|\hat{\kappa}_{\hat{\theta}_n} - \kappa_0\| = o_p(n^{-1/4})$, in which θ_0 and κ_0 are the true values of θ and κ , respectively, and the first derivatives of $\hat{\pi}(X_i, U_i; \theta, \kappa)$ with respect to θ and κ are continuous and bounded.

(C2) For some integers $G \geq 1$,

$$\max_i \left\{ E(e_i^{4G}|X_i, U_i), E(e_{t,i}^{4G}|X_i, U_i), |\mu_i|, |\mu_{t,i}| \right\} \leq \bar{C} < \infty, \quad a.s.$$

where $i = 1, \dots, n$, and \bar{C} is a positive constant.

(C3) For some integers $1 \leq G \leq \infty$,

$$M_n \xi_\pi^{-2G} \sum_{m=1}^{M_n} \{R_\pi(\omega_m^o)\}^G \xrightarrow{a.s.} 0.$$

(C4) $\bar{h} \xrightarrow{a.s.} 0$ and $(\bar{d} + \bar{k})\xi_\pi^{-2} \xrightarrow{a.s.} 0$.

(C5) The functions $\{g_j(U)\}_{j=1}^n$ belong to a class of functions \mathcal{F} , whose r th derivative $g_j^{(r)}$ exists and is Lipschitz of order η ,

$$\mathcal{F} = \left\{ g_j(\cdot) : |g_j^{(r)}(s) - g_j^{(r)}(t)| \leq G|s - t|^\eta \text{ for } s, t \in [a, b] \right\},$$

for some positive constant G , where r is a nonnegative integer and $\eta \in (0, 1]$, such that $\nu = r + \eta > 0.5$.

Condition (C1), a commonly used restriction for GPLMs, requires \sqrt{n} -consistent estimates for the parametric component $\hat{\theta}_n$. The nonparametric component $\hat{\kappa}_{\hat{\theta}_n}$ is viewed as a function of the parametric component to achieve consistency. This restriction is reasonable [20]. Condition (C2) is a moment condition concerning random errors, and it is satisfied by $\{\mu_i, \mu_{t,i}\}_{i=1}^n$, which are bounded. Condition (C3) is a convergence condition that imposes certain restrictions on the circumstances for applying our asymptotic outcome. A prerequisite for Condition (C3) to hold is $\xi_\pi \rightarrow \infty$, which requires that no finite-dimensional correct model exists in the class of candidate models [6]. It also requires that M_n and $\max_{1 \leq m \leq M_n} R_\pi(\omega_m^o)$ go to infinity slowly enough. Condition (C4), an assumption that excludes the case of extremely unbalanced design matrices as candidate models, is widely imposed in studies of optimal model averaging based on cross-validation, such as [7, 12], among others. The B-spline approximation in PLMs requires Condition (C5) with references to [4, 21], which is a regularity condition that necessitates the nonparametric coefficient function to be sufficiently smooth.

Theorem 1. *If Conditions (C1)–(C5) are satisfied, then*

$$\frac{L_{\hat{\pi}}(\hat{\omega}_{cv})}{\inf_{\omega \in \mathcal{H}_n} L_{\hat{\pi}}(\omega)} \xrightarrow{P} 1.$$

Theorem 1 demonstrates its asymptotic optimality by minimizing the approximate risk with squared error loss. This theorem illustrates that the square error with selected model weights $\hat{\omega}_{cv}$ minimized by the LOO-CV criterion is asymptotically equal to that of the infeasible optimal weight vector.

3.2. Weight convergence

In this subsection, we concentrate on the convergence properties of the optimal weights in model averaging. It should be noted that, in this article, the m th candidate model in (2.3) is deemed correctly specified or a correct model if there exists $\beta^{(m)*}$ such that $\mu = \mathbf{X}^{(m)}\beta^{(m)*} + \mathbf{g}(\mathbf{U})$; otherwise, model (2.3) is said to be misspecified or an incorrect model.

We first introduce some notation. \hat{s}_{cv} is defined as the sum of the weights assigned to the correct candidate models using our proposed method, which can be denoted as $\hat{s}_{cv} = \sum_{m=1}^{m_0} \hat{\omega}_{cv,m}$ in mathematical notation, where m_0 indicates that the first m_0 candidate models are all correctly specified. Let $\mathcal{H}_F = \{\omega \in [0, 1]^{M_n} : \sum_{m=m_0+1}^{M_n} \omega_m = 1\}$ be the weight set of all incorrect candidate models and $\xi_{\pi,F} = \inf_{\omega \in \mathcal{H}_F} R_{\pi}(\omega)$ is the optimal risk when the weights are assigned to all the misspecified candidate models. We specify some necessary conditions for further analysis based on $n \rightarrow \infty$.

(C6) For some integers $1 \leq G \leq \infty$,

$$\xi_{\pi,F}^{-2G} \max \left\{ m_0(\bar{d} + \bar{k})^{2G}, (M_n - m_0) \sum_{m=m_0}^{M_n-m_0} \{R_{\pi}(\omega_m^o)\}^G \right\} \xrightarrow{a.s.} 0.$$

(C7) $\bar{h} = O(n^{-1/2})$.

Condition (C6) requires $\xi_{\pi,F}^{-2G}$ to grow at a rate no slower than $m_0(\bar{d} + \bar{k})^{2G}$ and $(M_n - m_0) \sum_{m=m_0}^{M_n-m_0} \{R_{\pi}(\omega_m^o)\}^G$. It is worth noting that if $(M_n - m_0) \sum_{m=m_0}^{M_n-m_0} \{R_{\pi}(\omega_m^o)\}^G$ is larger than $m_0(\bar{d} + \bar{k})^{2G}$, then Condition (C6) is identical to $(M_n - m_0) \xi_{\pi,F}^{-2G} \sum_{m=m_0}^{M_n-m_0} \{R_{\pi}(\omega_m^o)\}^G \xrightarrow{a.s.} 0$ and is thus analogous to Condition (C3). Condition (C7), excluding the case of peculiar models as candidate models, is from [13].

Theorem 2. *If Conditions (C1)–(C7) are satisfied, then*

$$\hat{s}_{cv} \xrightarrow{P} 1.$$

Theorem 2 indicates that the model averaging estimator $\hat{\mu}(\hat{\omega}_{cv})$ of our proposal is sufficient for the sum of the weights assigned to the correct models to converge to one in probability as the sample size goes to infinity and automatically excludes the incorrect models.

4. Numerical studies

To demonstrate the theoretical properties of Section 3, in this section we conduct two Monte Carlo experiments on the finite-sample performance, where Case 1 verifies the asymptotic optimality of

Theorem 1 with the candidate models all misspecified, and Case 2 justifies the weight convergence property of Theorem 2 with at least one correct model specified. In addition, the superiority of our method is illustrated by applying it to the Diabetes dataset. For a better analysis, we also consider several relevant existing methods as competitors to our CPLJMA approach, including the model selection methods for AIC, BIC, and treatment effect cross-validation (TECV) proposed by [15]; Information criterion-based model averaging methods such as SAIC and SBIC; the equal weight method (EW); and treatment effect estimation by mixing (TEEM) and the Mallows averaging of partially linear models (MAPLM). We calculate the mean squared error (MSE) to assess the performances of the estimators, defined as $MSE = \frac{1}{nD} \sum_{d=1}^D \|\hat{\mu}(\omega)^{(d)} - \mu^{(d)}\|^2$, where $\mu^{(d)}$ and $\{\hat{\mu}(\omega)\}^{(d)}$ are the CATE and model averaging estimator in the d th replicate, respectively. D denotes the number of replicates of the simulation. Additionally, we calculated MSE_{median} in the empirical analysis, defined as $MSE_{median} = \text{median}_{d=1, \dots, D} MSE^{(d)}$, and the optimal rate, which is the percentage of the smallest MSE value. In complement to the rigorous numerical investigations detailed earlier, we embark on the task of determining the number of (interior) knots. Like the pivotal role of bandwidth in kernel smoothing, these knots, akin to tuning parameters, have a remarkable influence on the smoothness and adaptability of our spline models. The details of the numerical study and its results are described in the following subsections.

4.1. Monte Carlo study

Case 1: Without correct candidate models

The data-generating process (DGP) is as follows:

$$Y_{t,i} - Y_{c,i} = \mu_i + e_i = \sum_{j=1}^{500} X_{ij}\beta_j + g(U_i) + 0.5X_{i2}^2 + e_{t,i} - e_{c,i},$$

where $X_{i1} = 1$. $\{X_{ij}\}_{j=2}^J$, the covariates of the linear part, are generated from a multivariate normal distribution with mean 0 and covariance $0.5^{|j_1-j_2|}$ between x_{ij_1} and x_{ij_2} . The associated coefficients in the linear component are taken as $\beta_j = 1/j$. The coefficient function $g(U_i) = \sin(2\pi U_i)$, where $U_i \sim \text{Uniform}[0, 1]$. $\{e_{t,i}, e_{c,i}\}$ are independent random errors distributed from $N(0, \sigma^2 X_{i2}^2)$, where the parameter σ^2 is chosen by $R^2 = \text{var}(\mu_i)/\text{var}(Y_{t,i} - Y_{c,i})$, which varies on a grid between 0.1 and 0.9. Then

$$\mu_i = \sum_{j=1}^{500} j^{-1} X_{ij} + \sin(2\pi U_i) + 0.5X_{i2}^2 \quad \text{and} \quad e_i = e_{t,i} - e_{c,i}. \quad (4.1)$$

We rescaled μ_i to have unit variance so that the expected R^2 equals $\frac{1}{1+\sigma^2}$ for the unknown model. It is clear from (4.1) that the class of candidate models considered in this case is misspecified. In addition, to obtain $\{\delta\}_{i=1}^n$, the propensity score is taken as

$$\pi(X_i, U_i) = \frac{\exp(0.75X_{i2} + \sin(2\pi U_i))}{1 + \exp(0.75X_{i2} + \sin(2\pi U_i))}.$$

Thus, we obtain $\{Y_i\}_{i=1}^n$. For $\hat{\pi}(X_i, U_i)$, we use the GPLMs in (2.6) to approximate the coefficients of the linear part and the form of the nonparametric part.

First, we delve into the influence of interior knots within the B-spline basis on the performance of our proposed CPLJMA approach. By employing the “ $bs(\cdot, df)$ ” function from the **R** package “*splines*”, we generate an appropriate B-spline basis matrix. Here, the degree of freedom parameter, denoted as df , is a crucial factor, determined by “ $df = 3 + \text{the number of knots}$ ”.

Figure 1 shows the variation in MSE as the number of interior knots varies, considering a sample size of $n = 300$ and $R^2 = 0.5$, and for both nested and nonnested setting scenarios. In the nested setting, the m th candidate model comprises the first m linear variables in $\{X_{ij}\}_{j=1}^{500}$. The number of candidate models M_n is determined as the nearest integer from $3n^{1/3}$, resulting in $M_n = 20$. In the nonnested setting, the linear components of all candidate models are a subset of $\{X_{i1}, \dots, X_{i5}\}$, disregarding the remaining X_i variables, thereby yielding a total of $2^5 - 1 = 31$ candidate models. As depicted in the figure, the MSE increases with an increase in the number of knots, potentially exacerbating the phenomenon of overfitting. Consequently, we opt for “ $df = 3$ ” as the degree of freedom for the CPLJMA method, ensuring a balance between model flexibility and susceptibility to overfitting.

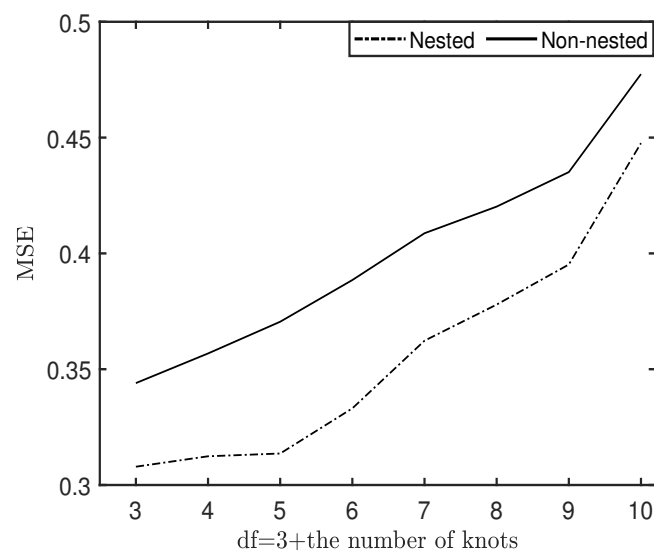


Figure 1. The curves of the median of MSE with the number of knots for Case 1 over $n = 300$ and $R = 0.5$ simulations.

The greater the number of covariates, the more tedious the computation. Therefore, in substantiating the theoretical properties outlined in Theorem 1 through numerical simulations, we adopt a nested setting. Accordingly, when the sample sizes are $n = 75, 150, 300,$ and 600 , then $M_n = 12, 15, 20,$ and 25 , respectively.

Figure 2 illustrates a numerical inspection for the asymptotic optimality of $\hat{\mu}_{\hat{\pi}}(\hat{\omega}_{cv})$ in Theorem 1 by showing the mean of LR, defined as $LR = L_{\hat{\pi}}(\hat{\omega}_{cv}) / \inf_{\omega \in \mathcal{H}_n} L_{\hat{\pi}}(\omega)$, for different samples and various R . In a simulation trial based on 100 replications, we observe that the mean curve of LR decreases and converges to one as n increases. This intuitively demonstrates the asymptotic optimality of CPLJMA.

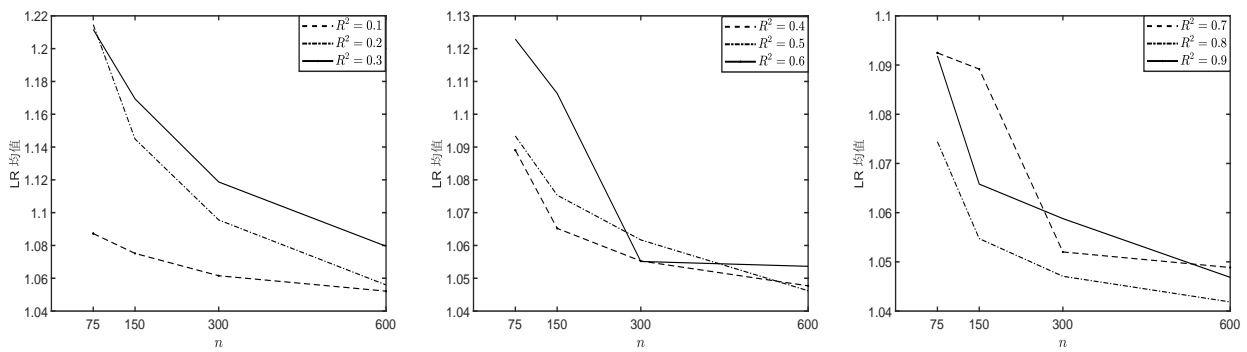


Figure 2. Evaluating the asymptotic optimality of $\hat{\mu}_{\hat{\pi}}(\hat{\omega}_{cv})$ under nested models: The class of candidate models is all misspecified.

Figure 3 shows the MSE ratio curves for the seven CATE μ estimators we considered, where we used the AIC as the denominator to yield the MSE ratio with an entry of 1.00. Generally, our proposed CPLJMA outperforms its competitors on the MSE ratios when R^2 or n is small or moderate, particularly because it is difficult to identify the optimal model when there is considerable noise in the model. The advantage of model averaging without relying on a single model is that it provides protection against poor model selection. As expected, SAIC and SBIC invariably yielded more accurate results than did their respective model selection rivals. In short, to some extent, CPLJMA is superior to its competitors.

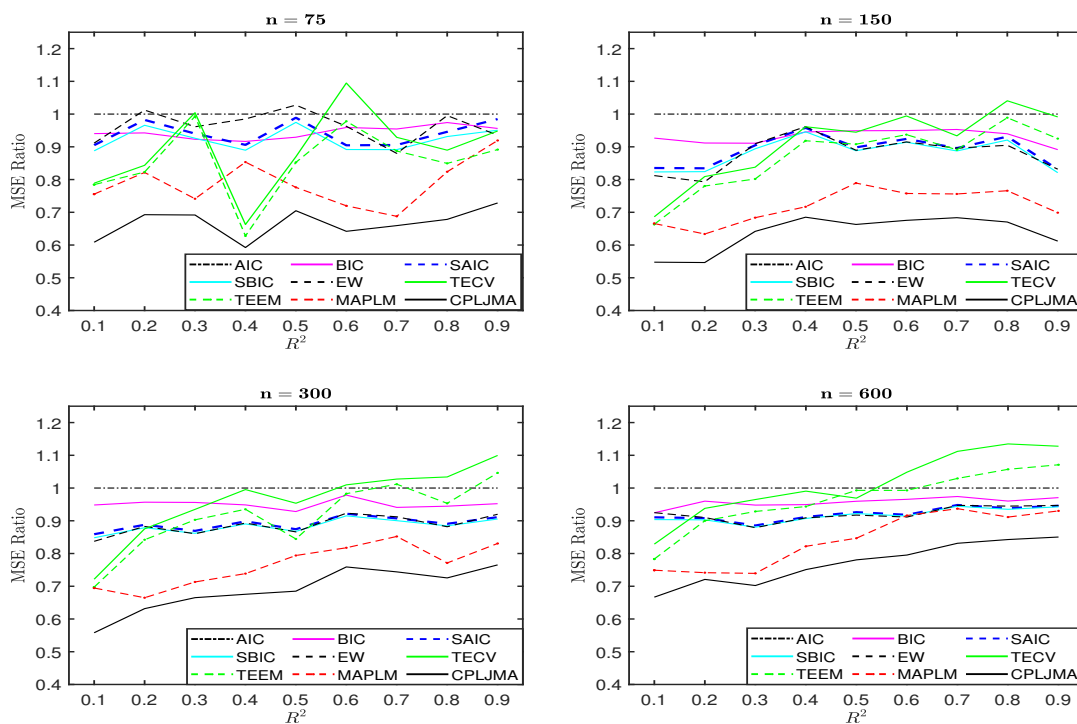


Figure 3. Performances of various method estimators under nested models: The classes of candidate models are all misspecified.

Case 2: With correct candidate models

The DGP is generated from

$$Y_{t,i} - Y_{c,i} = \mu_i + e_i = \sum_{j=1}^5 X_{ij}\beta_j + g(U_i) + e_{t,i} - e_{c,i}$$

where the vector of covariates $X_i = (X_{i1}, \dots, X_{i5})^T$ is from an independent standard normal distribution $N(0, 1)$, U_i is distributed as $U[-1, 1]$, and the corresponding coefficient and nonparametric function are $\beta_j = 1/j$ and $1.2U_i$, respectively. The settings for $\{e_{t,i}, e_{c,i}\}$, σ , R^2 , and bf are the same as those in Case 1. Thus, we can obtain

$$\mu_i = \sum_{j=1}^5 j^{-1} X_{ij} + 1.2U_i.$$

The propensity score is taken as

$$\pi(X_i, U_i) = \frac{\exp(0.75X_{i3} + 1.2U_i)}{1 + \exp(0.75X_{i3} + 1.2U_i)}.$$

Based on the LPLMs in (2.6), we can obtain $\{\hat{\pi}(X_i, U_i)\}_{i=1}^n$ and thus $\{Z_{\pi,i}\}_{i=1}^n$. In this case, we consider the nonnested models. The linear parts of all candidate models are constructed by varying combinations of $\{X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5}\}$; thus, $M_n = 2^5 - 1$. The sample size is taken as $n = 75, 150, 300, 600$. The results for the convergence of the model weights $\hat{\omega}_{cv}$ and the MSE ratio of the above methods are given in Figures 4 and 5, respectively, based on 100 replications.

Figure 4 clearly shows that the sum of the weights corresponding to the correct candidate models tends to one as the sample sizes n and R^2 increase. This intuitively confirms the convergence of $\hat{\omega}_{cv}$ presented in Theorem 2 via numerical inspection.

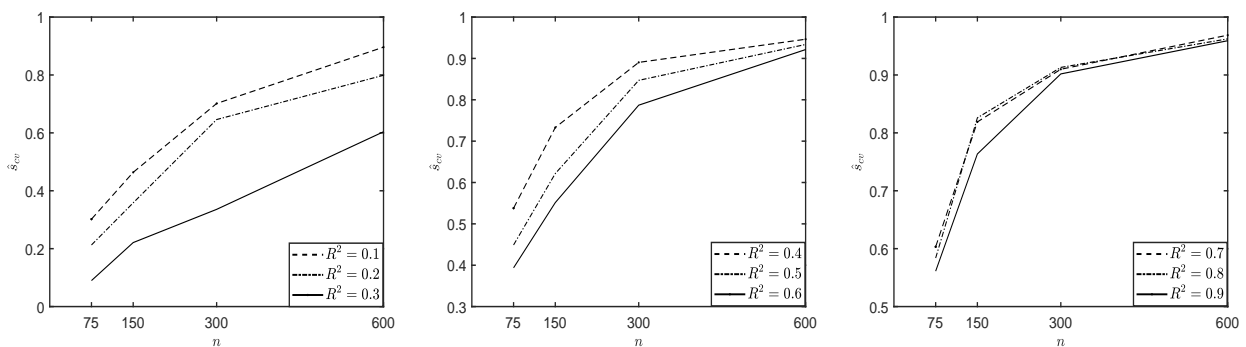


Figure 4. Evaluating the convergence of the model weights $\hat{\omega}_{cv}$ under nonnested models: The class of candidate models contains correct models.

As shown in Figure 5, we still derive the MSE ratio using AIC as the denominator. In most cases, the results of the MSE ratio demonstrate the merits of our approach over its competitors. Obviously, as R^2 increases progressively, the MSE ratio for all scenarios tends to decrease, as expected. Increasing the sample size also improved the performance for all approaches. Overall, CPLJMA still outperforms several existing methods.

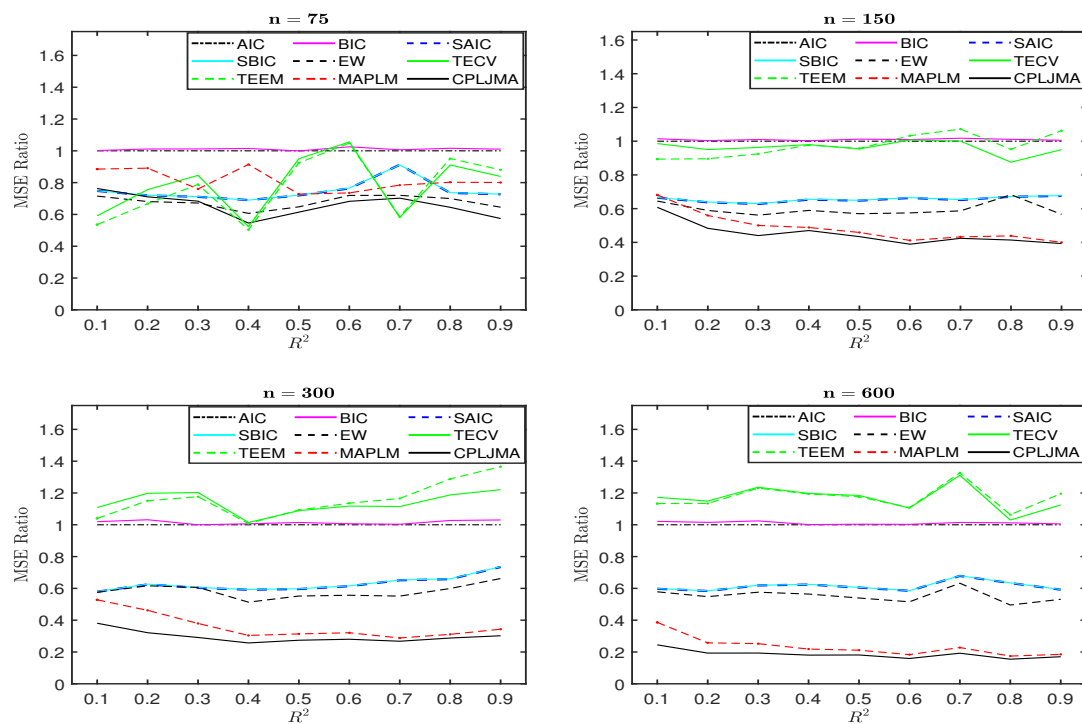


Figure 5. MSE ratio curves for the various methods under nonnested models: The class of candidate models contains correct models.

4.2. Empirical analysis

Our proposed method is applied to the Diabetes dataset from Dr. John Schorling, Department of Medicine, University of Virginia School of Medicine. The original data consisted of 19 covariates on 403 subjects from 1046 subjects interviewed in a study. However, due to the availability of missing data, we selected 16 covariates and 366 respondents as the dataset for the current case study, 175 of whom respondents resided in Buckingham and 191 of whom did not.

Our analysis considers the outcome variable \mathbf{Y} to be stabilized glucose. The treatment indicator variable δ takes the value of 1 if people reside in Buckingham, and 0 otherwise. We calculated the Pearson correlation coefficients of the 14 baseline covariates \mathbf{X} and \mathbf{U} with \mathbf{Y} and ranked them in descending order of correlation strength. Let us note that all continuous covariates \mathbf{X} and \mathbf{Y} are standardized to have a mean of zero and a variance of one, and \mathbf{U} is scaled to $[0, 1]$. See Table 1 for details.

Table 1. Pearson correlation coefficient of each linear covariate with \mathbf{Y} .

Symbol	Description	Correlation with \mathbf{Y}
X_1	glycosolated hemoglobin	0.7409
X_2	cholesterol/HDL ratio	0.2989
X_3	waist	0.2337
X_4	weight	0.1888
X_5	high density lipoprotein	-0.1801
X_6	frame (0 if large, 1 if medium, 2 otherwise)	-0.1726
X_7	first systolic blood pressure	0.1654
X_8	total cholesterol	0.1514
X_9	hip	0.1448
X_{10}	gender (0 if male, 1 if otherwise)	-0.0861
X_{11}	height	0.0825
X_{12}	postprandial time when labs were drawn	-0.0485
X_{13}	first diastolic blood pressure	0.0257

We assume that the candidate models consist of nested models constructed from the covariates in $\{X_1, \dots, X_{13}, \mathbf{U}\}$ with intercept terms, where the baseline covariates are \mathbf{X} for the linear part, and we consider that age is a covariate \mathbf{U} for the nonparametric part. Accordingly, there are 13 well-prepared candidate models. To implement our proposal, the propensity score $\pi(X_i, U_i)$ is still solved by LPLMs.

We conduct a “guided simulation experiment” to evaluate the performance of our proposal and that of its competitors. In particular, we use the largest candidate model containing all covariates as a guided model, and m^* denotes the index of that model in the class of candidate models. Based on the m^* th candidate model and the original dataset $\{Y_i, X_i, U_i, \delta_i\}_{i=1}^n$, we can obtain a simulation dataset $\{Y_i^{(m^*)}, X_i, \delta_i, U_i\}_{i=1}^n$. Thus,

$$\begin{aligned}
 Y_i^{(m^*)} &= \delta_i Y_{t,i}^{(m^*)} + (1 - \delta_i) Y_{c,i}^{(m^*)}, \\
 Y_{t,i}^{(m^*)} &= X_i^{(m^*)T} \hat{\rho}^{(m^*)} + f^{(m^*)}(U_i^{(m^*)}) + e_{t,i}^{(m^*)}, \\
 Y_{c,i}^{(m^*)} &= X_i^{(m^*)T} \hat{\eta}^{(m^*)} + h^{(m^*)}(U_i^{(m^*)}) + e_{c,i}^{(m^*)},
 \end{aligned} \tag{4.2}$$

where $\hat{\rho}^{(m^*)}$ and $\hat{\eta}^{(m^*)}$ are the regression coefficient estimators for the linear part, $f^{(m^*)}(U_i^{(m^*)})$ and $h^{(m^*)}(U_i^{(m^*)})$ are the estimators for the nonparametric part, and $\{e_{t,i}^{(m^*)}, e_{c,i}^{(m^*)}\}_{i=1}^n$ is from $N(0, 1)$. Therefore, the “true” μ_i , CATE, is known in this analysis dataset, namely,

$$\mu_i = X_i^{(m^*)T} \hat{\rho}^{(m^*)} + f^{(m^*)}(U_i^{(m^*)}) - [X_i^{(m^*)T} \hat{\eta}^{(m^*)} + h^{(m^*)}(U_i^{(m^*)})].$$

We randomly selected samples from 20%, 40%, 60%, 80%, and 100% of the dataset to describe the performance of the proposed CPLJMA and its competitors by MSE, $MS E_{median}$, and the optimal rate based on 100 replications.

The results are displayed in Table 2. Our approach produced a lower MSE and median and a higher optimal rate than those of its competitors across all sample sizes considered. As expected, the average methods based on the information criterion perform better than the model selection methods. To some extent, our proposal has a clear advantage over its competitors in solving practical problems.

Table 2. The MSE, Median and the Optimal rate across 100 repetitions under nested models.

n	Method	AIC	BIC	SAIC	SBIC	EW	TECV	TEEM	MAPLM	CPLJMA
20%	MSE	2.0044	1.9531	1.8427	1.8182	1.9216	1.9145	1.8024	1.8539	1.5563
	Median	0.7534	0.7321	0.6935	0.6824	0.7322	0.7035	0.5806	0.6822	0.5255
	Optimal rate	0.13	0.04	0.01	0.11	0.05	0.02	0.11	0.13	0.40
40%	MSE	1.1867	1.6152	1.0405	1.0329	0.9579	0.9649	0.8924	0.9416	0.8520
	Median	0.4276	0.4165	0.3666	0.3624	0.3522	0.3435	0.2892	0.3237	0.2876
	Optimal rate	0.13	0.00	0.02	0.07	0.01	0.04	0.15	0.18	0.40
60%	MSE	1.0056	0.9915	0.8794	0.8749	0.9020	0.8820	0.8467	0.8835	0.7948
	Median	0.3621	0.3549	0.3106	0.3078	0.3243	0.3020	0.2557	0.3069	0.2598
	Optimal rate	0.06	0.04	0.02	0.07	0.03	0.04	0.15	0.14	0.45
80%	MSE	0.8512	0.8336	0.7542	0.7513	0.7522	0.7500	0.7219	0.7282	0.6955
	Median	0.3055	0.2970	0.2600	0.2591	0.2576	0.2479	0.2306	0.2392	0.2272
	Optimal rate	0.05	0.00	0.00	0.11	0.02	0.04	0.24	0.16	0.38
100%	MSE	0.8357	0.8247	0.7105	0.7082	0.7013	0.7173	0.7669	0.6824	0.6628
	Median	0.2997	0.2935	0.2478	0.2465	0.2426	0.2375	0.2667	0.2283	0.2184
	Optimal rate	0.06	0.00	0.01	0.03	0.05	0.13	0.01	0.12	0.59

5. Conclusions

Considering the heterogeneity and heteroskedasticity information embedded in the dataset, the problem of model uncertainty, and the flexibility and interpretability of PLMs, a jackknife-type method based on the weight selection criterion and its feasible form, called the CPLJMA method, are proposed for estimating the CATE. Within this context, we consider the optimal model weights chosen by minimizing the LOO-CV criterion, in which the B-splines approximate the nonparametric function, and we demonstrate its asymptotic optimality in accordance with a minimization of the approximate risk with squared error loss. In addition, since the choice of weights is crucial to the model averaging method, we considered the convergence properties of the weights, that is, when the sample size goes to infinity and at least one candidate model is correctly specified, the sum of weights obtained by our approach for the correct candidate models converges to one in probability. In the simulation section, we examine the finite-sample performance of our estimator and compare it with several other model selection and averaging methods. We illustrate our method by using a real-world dataset. The simulation results indicated that our method possessed some advantages relative to its competitors.

There are still many issues worthy of further discussion. First, our proposed CPLJMA method is valid only for one-dimensional covariates with respect to the nonparametric part, and it can be further refined and extended to multiple dimensions. Second, the least squares loss used as the loss function in our analysis is more sensitive to outliers. Thus, quantile regression, which is less sensitive to outliers, is considered as a loss function because developing the model averaging process and establishing its asymptotic properties remains challenging; this area deserves future research.

Author contributions

X. Zhang: Writing—original draft; X. Zhang and J. Li: Writing—review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank the editors and reviewers for their valuable and insightful comments. The authors also thank Professor Guangren Yang for his advice for revising the paper.

Conflict of interest

The authors declare no conflict of interest.

References

1. B. A. Brumback, *Fundamentals of causal inference: With R*, CRC Press, 2021. <https://doi.org/10.1080/01621459.2023.2287599>
2. R. K. Crump, V. J. Hotz, G. W. Imbens, O. A. Mitnik, Nonparametric tests for treatment effect heterogeneity, *Rev. Econ. Stat.*, **90** (2008), 389–405. <https://doi.org/10.1162/rest.90.3.389>
3. R. F. Engle, C. W. J. Granger, J. Rice, A. Weiss, Semiparametric estimates of the relation between weather and electricity sales, *J. Am. Stat. Assoc.*, **81** (1986), 247–269. <https://doi.org/10.1080/01621459.1986.10478274>
4. J. Fan, Y. Ma, W. Dai, Nonparametric independence screening in sparse ultra-high-dimensional varying coefficient models, *J. Am. Stat. Assoc.*, **109** (2014), 1270–1284. <https://doi.org/10.1080/01621459.2013.879828>
5. Y. Gao, W. Long, Z. Wang, Estimating average treatment effect by model averaging, *Econ. Lett.*, **135** (2015), 42–45. <https://doi.org/10.1016/j.econlet.2015.08.002>
6. B. E. Hansen, Least squares model averaging, *Econometrica*, **75** (2007), 1175–1189. <https://doi.org/10.1111/j.1468-0262.2007.00785.x>
7. B. E. Hansen, J. S. Racine, Jackknife model averaging, *J. Econ.*, **167** (2012), 38–46. <https://doi.org/10.1016/j.jeconom.2011.06.019>
8. G. W. Imbens, J. M. Wooldridge, Recent developments in the econometrics of program evaluation, *J. Econ. Lit.*, **47** (2009), 5–86. <https://doi.org/10.1257/jel.47.1.5>
9. K. Imai, M. Ratkovic, Estimating treatment effect heterogeneity in randomized program evaluation, *Ann. Appl. Stat.*, **7** (2013), 443–470. <https://doi.org/10.1214/12-AOAS593>

10. H. Jo, M. A. Harjoto, The causal effect of corporate governance on corporate social responsibility, *J. Bus. Ethics.*, **106** (2012), 53–72. <https://doi.org/10.1007/s10551-011-1052-1>
11. T. Kitagawa, C. Muris, Model averaging in semiparametric estimation of treatment effects, *J. Econ.*, **193** (2016), 271–289. <https://doi.org/10.1016/j.jeconom.2016.03.002>
12. K. C. Li, Asymptotic optimality for C_p , C_L , cross-validation and generalized cross-validation: Discrete index set, *Ann. Stat.*, **15** (1987), 958–975. <https://doi.org/10.1214/aos/1176350486>
13. Q. Liu, R. Okui, Heteroskedasticity-robust C_p model averaging, *Econom. J.*, **16** (2013), 463–472. <https://doi.org/10.1111/ectj.12009>
14. M. Müller, Estimation and testing in generalized partial linear models—A comparative study, *Stat. Comput.*, **11** (2001), 299–309. <https://doi.org/10.1023/A:1011981314532>
15. C. A. Rolling, Y. Yang, Model selection for estimating treatment effects, *J. R. Stat. Soc. B*, **76** (2014), 749–769. <https://doi.org/10.1111/rssb.12043>
16. C. A. Rolling, Y. Yang, D. Velez, Combining estimates of conditional treatment effects, *Economet. Theor.*, **35** (2019), 1089–1110. <https://doi.org/10.1017/S0266466618000397>
17. P. R. Rosenbaum, D. B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70** (1983), 41–55. <https://doi.org/10.2307/2335942>
18. D. B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, *J. Educ. Psychol.*, **66** (1974), 688–701. <https://doi.org/10.1037/h0037350>
19. D. B. Rubin, Assignment to treatment group on the basis of a covariate, *J. Educ. Behav. Stat.*, **2** (1977), 1–26. <https://doi.org/10.2307/1164933>
20. T. A. Severini, J. G. Staniswalis, Quasi-likelihood estimation in semiparametric models, *J. Am. Stat. Assoc.*, **89** (1994), 501–511. <https://doi.org/10.2307/2290852>
21. C. J. Stone, Optimal global rates of convergence for nonparametric regression, *Ann. Stat.*, **10** (1982), 1040–1053. <https://doi.org/10.1214/aos/1176345969>
22. L. Tian, A. A. Alizadeh, A. J. Gentles, R. Tibshirani, A simple method for estimating interactions between a treatment and a large number of covariates, *J. Am. Stat. Assoc.*, **109** (2014), 1517–1532. <https://doi.org/10.1080/01621459.2014.951443>
23. W. Whittle, Bounds for the moments of linear and quadratic forms in independent variables, *Theory Probab. Appl.*, **5** (1960), 302–305. <https://doi.org/10.1137/1105028>
24. Z. Tan, On doubly robust estimation for logistic partially linear models, *Stat. Probab. Lett.*, **155** (2019), 108577. <https://doi.org/10.1016/j.spl.2019.108577>
25. J. Zeng, W. Cheng, G. Hu, Y. Rong, Model selection and model averaging for semiparametric partially linear models with missing data, *Commun. Stat.-Theor. M.*, **48** (2019), 381–395. <https://doi.org/10.1080/03610926.2017.1410717>
26. X. Zhang, A. T. Wan, G. Zou, Model averaging by jackknife criterion in models with dependent data, *J. Econ.*, **174** (2013), 82–94. <https://doi.org/10.1016/j.jeconom.2013.01.004>
27. X. Zhang, W. Wang, Optimal model averaging estimation for partially linear models, *Stat. Sin.*, **29** (2019), 693–718. <https://doi.org/10.2139/ssrn.2948380>

Supplementary

This Supplementary provides detailed proofs of the main theorems stated in the above.

Lemmas and their proofs

We introduce some lemmas and their proofs before proving the Theorems in Section 3.

Lemma 1. Provided that Conditions (C1)-(C2) hold, we have $\|Z_{\hat{\pi}} - Z_{\pi}\|^2 = O_p(1)$.

Proof. According to the monotonicity of the L_r -norm and the Holder inequality, we have

$$\begin{aligned} \max \left\{ \max_{1 \leq i \leq n} \sigma_i^2, \max_{1 \leq i \leq n} \sigma_{t,i}^2 \right\} &= \max \left\{ \max_{1 \leq i \leq n} E(e_i^2 | X_i, U_i), \max_{1 \leq i \leq n} E(e_{t,i}^2 | X_i, U_i) \right\} \\ &\leq \max \left[\max_{1 \leq i \leq n} \{E(e_i^{4G} | X_i, U_i)\}^{\frac{1}{2G}}, \max_{1 \leq i \leq n} \{E(e_{t,i}^{4G} | X_i, U_i)\}^{\frac{1}{2G}} \right] \\ &\leq \bar{C} \end{aligned}$$

almost surely, where the second inequality and the last inequality are attributed to the Holder inequality and Condition (C2), respectively. Thus, we know that for any $\epsilon > 0$, there exists an integer N_ϵ such that $P(\max_{1 \leq i \leq n} \sigma_i^2 > \bar{C}) \leq \epsilon/2$ for all $n \geq N_\epsilon$. Let $M_\epsilon = 2\bar{C}/\epsilon$. Then we have

$$\begin{aligned} &\sup_{n \geq 1} P \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon \right) \\ &= \sup_{n \geq 1} \left\{ P \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon, \max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) + P \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon, \max_{1 \leq i \leq n} \sigma_i^2 > \bar{C} \right) \right\} \\ &\leq \sup_{n \geq 1} E \left\{ I \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon \right) I \left(\max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) \right\} + \sup_{n \geq N_\epsilon} P \left(\max_{1 \leq i \leq n} \sigma_i^2 > \bar{C} \right) \\ &\leq \sup_{n \geq 1} E \left[E \left\{ I \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon \right) \middle| X_i, U_i \right\} I \left(\max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) \right] + \frac{\epsilon}{2} \\ &= \sup_{n \geq 1} E \left\{ P \left(\frac{1}{n} \sum_{i=1}^n e_i^2 > M_\epsilon \middle| X_i, U_i \right) I \left(\max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) \right\} + \frac{\epsilon}{2} \\ &\leq \sup_{n \geq 1} E \left\{ M_\epsilon^{-1} E \left(\frac{1}{n} \sum_{i=1}^n e_i^2 \middle| X_i, U_i \right) I \left(\max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) \right\} + \frac{\epsilon}{2} \\ &= \sup_{n \geq 1} E \left\{ M_\epsilon^{-1} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 I \left(\max_{1 \leq i \leq n} \sigma_i^2 \leq \bar{C} \right) \right\} + \frac{\epsilon}{2} \\ &\leq M_\epsilon^{-1} \bar{C} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned} \tag{5.1}$$

Thus, we have

$$\frac{1}{n} \sum_{i=1}^n e_i^2 = O_p(1). \tag{5.2}$$

Similarly, it can be obtained that

$$\frac{1}{n} \sum_{i=1}^n e_{t,i}^2 = O_p(1). \quad (5.3)$$

By the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \|Z_{\hat{\pi}} - Z_{\pi}\|^2 \\ &= \sum_{i=1}^n \left\{ \frac{\delta_i}{\hat{\pi}(X_i, U_i)} Y_{t,i} - \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} Y_{c,i} - \left(\frac{\delta_i}{\pi(X_i, U_i)} Y_{t,i} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} Y_{c,i} \right) \right\}^2 \\ &= \sum_{i=1}^n \left[\left\{ \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right\} Y_{t,i} \right. \\ &\quad \left. + \left\{ \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} \right\} (Y_{t,i} - Y_{c,i}) \right]^2 \\ &\leq c \left[\left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right| \right\}^2 \frac{1}{n} \sum_{i=1}^n Y_{t,i}^2 \right. \\ &\quad \left. + \left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} \right| \right\}^2 \frac{1}{n} \sum_{i=1}^n (Y_{t,i} - Y_{c,i})^2 \right] \\ &\leq c \left[\left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right| \right\}^2 \left(\frac{1}{n} \sum_{i=1}^n \mu_{t,i}^2 + \frac{1}{n} \sum_{i=1}^n e_{t,i}^2 \right) \right. \\ &\quad \left. + \left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} \right| \right\}^2 \left(\frac{1}{n} \sum_{i=1}^n \mu_i^2 + \frac{1}{n} \sum_{i=1}^n e_i^2 \right) \right] \\ &\leq \left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right| \right\}^2 O_p(1) \\ &\quad + \left\{ \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} \right| \right\}^2 O_p(1), \end{aligned}$$

where the last equation is due to Condition (C2), (5.2) and (5.3). Lemma 1 holds, if we can prove that

$$\sqrt{n} \max_{1 \leq i \leq n} \left| \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right| = O_p(1), \quad (5.4)$$

$$\sqrt{n} \max_{1 \leq i \leq n} \left| \frac{1 - \delta_i}{1 - \hat{\pi}(X_i, U_i)} - \frac{1 - \delta_i}{1 - \pi(X_i, U_i)} \right| = O_p(1). \quad (5.5)$$

By the Taylor expansion, one has

$$\begin{aligned} & \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{\delta_i - \hat{\pi}(X_i, U_i)}{\hat{\pi}(X_i, U_i)(1 - \hat{\pi}(X_i, U_i))} - \frac{\delta_i - \pi(X_i, U_i)}{\pi(X_i, U_i)(1 - \pi(X_i, U_i))} \right| \\ &\leq c \left\{ \min_{1 \leq i \leq n} \hat{\pi}(X_i, U_i; \theta_i^*, \kappa_i^*) \right\}^{-2} \left\{ 1 - \max_{1 \leq i \leq n} \hat{\pi}(X_i, U_i; \theta_i^*, \kappa_i^*) \right\}^{-2} \\ &\quad \cdot \max_{1 \leq i \leq n} \left\{ \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \theta^T} \right|_{\theta=\theta_i^*} \right\| \sqrt{n} \|\hat{\theta}_n - \theta_0\| + \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \kappa^T} \right|_{\kappa=\kappa_i^*} \right\| \|\kappa_{\hat{\theta}_n} - \kappa_0\| \Big\}, \end{aligned}$$

where θ_i^* is a vector between $\hat{\theta}_n$ and θ_0 , and κ_i^* is a vector between $\kappa_{\hat{\theta}_n}$ and κ_0 . Expanding $\hat{\pi}(X_i, U_i; \theta_i^*, \kappa_i^*)$ in a Taylor series and considering the property of $\pi(X_i, U_i)$, we have

$$\min_{1 \leq i \leq n} \hat{\pi}(X_i, U_i; \theta_i^*, \kappa_i^*) \geq c_\pi - \max_{1 \leq i \leq n} \left\{ \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \theta^T} \right\|_{\theta=\theta_i^{**}} \|\hat{\theta}_n - \theta_0\| + \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \kappa^T} \right\|_{\kappa=\kappa_i^{**}} \|\kappa_{\hat{\theta}_n} - \kappa_0\| \right\},$$

$$1 - \max_{1 \leq i \leq n} \hat{\pi}(X_i, U_i; \theta_i^*, \kappa_i^*) \geq c_\pi - \max_{1 \leq i \leq n} \left\{ \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \theta^T} \right\|_{\theta=\theta_i^{**}} \|\hat{\theta}_n - \theta_0\| + \left\| \frac{\partial \hat{\pi}(X_i, U_i; \theta_i, \kappa_i)}{\partial \kappa^T} \right\|_{\kappa=\kappa_i^{**}} \|\kappa_{\hat{\theta}_n} - \kappa_0\| \right\},$$

where θ_i^{**} is a vector between $\hat{\theta}_i^*$ and θ_0 , and κ_i^{**} is a vector between $\hat{\kappa}_i^*$ and κ_0 . Together with Condition (C2), this shows that (5.4) is valid. Likewise, we determine that (5.5) is valid. Therefore, the proof of Lemma 1 is completed. \square

Lemma 2. By Condition (10) of Theorem 2.1 in [26], we have

$$\sup_{\omega \in \mathcal{H}_n} \left| \frac{\tilde{R}_\pi(\omega)}{R_\pi(\omega)} - 1 \right| \xrightarrow{a.s.} 0. \quad (5.6)$$

With Assumption 3, for some integer $1 \leq G \leq \infty$, there exists

$$M_n \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \{\tilde{R}_\pi(\omega_m^o)\}^G \xrightarrow{a.s.} 0. \quad (5.7)$$

Lemma 3. Let $\tilde{\omega} = \operatorname{argmin}_{\omega \in \mathcal{H}_n} \{L_n(\omega) + a_n(\omega) + b_n\}$, where $a_n(\omega)$ is a term related to ω and b_n is a term unrelated to ω . Let $R_n(\omega) = E\{L_n(\omega)|X, U\}$. If

$$\sup_{\omega \in \mathcal{H}_n} \frac{|a_n(\omega)|}{R_n(\omega)} = o_p(1), \quad (5.8)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{|R_n(\omega) - L_n(\omega)|}{R_n(\omega)} = o_p(1), \quad (5.9)$$

and there exists a constant c and a positive integer N^* so that when $n \geq N^*$, $\inf_{\omega \in \mathcal{H}_n} R_n(\omega) \geq c > 0$ almost surely, then $L_n(\tilde{\omega}) / \inf_{\omega \in \mathcal{H}_n} L_n(\omega) \rightarrow 1$ in probability.

Lemma 4. For any $n_1 \times n_2$ matrices \mathbf{B}_1 and \mathbf{B}_2 ,

$$\lambda_{\max}\{\mathbf{B}_1 \mathbf{B}_2\} \leq \lambda_{\max}\{\mathbf{B}_1\} \lambda_{\max}\{\mathbf{B}_2\}, \quad \lambda_{\max}\{\mathbf{B}_1 + \mathbf{B}_2\} \leq \lambda_{\max}\{\mathbf{B}_1\} + \lambda_{\max}\{\mathbf{B}_2\}.$$

Theorems and theirs proofs

Proof of Theorem 1:

$\lambda_{\max}(\cdot)$ is denoted as the largest singular value of a matrix. From Condition (C2), we have

$$\lambda_{\max}(\mathbf{\Omega}_\pi) = O(1). \quad (5.10)$$

By the definition of $R_\pi(\omega)$, it can be shown that

$$R_\pi(\omega) = \|\mathbf{A}(\omega)\boldsymbol{\mu}\|^2 + \operatorname{tr}\{\mathbf{P}(\omega)\mathbf{\Omega}_\pi\mathbf{P}(\omega)^T\}, \quad (5.11)$$

where $\mathbf{A}(\omega) = \mathbf{I}_n - \mathbf{P}(\omega)$. Define the loss function of $\tilde{\boldsymbol{\mu}}_\pi(\omega)$ as $\tilde{L}_\pi(\omega) = \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2$ and its corresponding risk function as $\tilde{R}_\pi(\omega) = E\{\tilde{L}_\pi(\omega)|X, U\}$. Similarly, we have $\tilde{R}_\pi(\omega) = \|\tilde{\mathbf{A}}(\omega)\boldsymbol{\mu}\|^2 + \text{tr}\{\tilde{\mathbf{P}}(\omega)\boldsymbol{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^T\}$ in which $\tilde{\mathbf{A}}(\omega) = \mathbf{I}_n - \tilde{\mathbf{P}}(\omega)$. Define

$$V_\pi(\omega) = \|\mathbf{A}(\omega)\boldsymbol{\mu}\|^2 + \text{tr}\{\mathbf{P}(\omega)\boldsymbol{\Omega}_\pi\mathbf{P}(\omega)^T\} \quad \text{and} \quad \tilde{V}_\pi(\omega) = \|\tilde{\mathbf{A}}(\omega)\boldsymbol{\mu}\|^2 + \text{tr}\{\tilde{\mathbf{P}}(\omega)\boldsymbol{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^T\}.$$

Then we have

$$\begin{aligned} \frac{L_{\hat{\pi}}(\hat{\omega}_{cv})}{\inf_{\omega \in \mathcal{H}_n} L_{\hat{\pi}}(\omega)} - 1 &= \sup_{\omega \in \mathcal{H}_n} \left\{ \frac{L_{\hat{\pi}}(\hat{\omega}_{cv})}{L_{\hat{\pi}}(\omega)} - 1 \right\} \\ &= \sup_{\omega \in \mathcal{H}_n} \left\{ \frac{L_{\hat{\pi}}(\hat{\omega}_{cv})}{V_\pi(\hat{\omega}_{cv})} \frac{V_\pi(\hat{\omega}_{cv})}{\tilde{V}_\pi(\hat{\omega}_{cv})} \frac{\tilde{V}_\pi(\hat{\omega}_{cv})}{\tilde{L}_\pi(\hat{\omega}_{cv})} \frac{\tilde{L}_\pi(\hat{\omega}_{cv})}{\tilde{L}_\pi(\omega)} \frac{\tilde{R}_\pi(\omega)}{\tilde{R}_\pi(\omega)} \frac{R_\pi(\omega)}{L_{\hat{\pi}}(\omega)} - 1 \right\} \\ &\leq \sup_{\omega \in \mathcal{H}_n} \left(\frac{L_{\hat{\pi}}(\omega)}{R_\pi(\omega)} \right) \sup_{\omega \in \mathcal{H}_n} \left(\frac{R_\pi(\omega)}{\tilde{R}_\pi(\omega)} \right) \sup_{\omega \in \mathcal{H}_n} \left(\frac{\tilde{R}_\pi(\omega)}{\tilde{L}_\pi(\omega)} \right) \sup_{\omega \in \mathcal{H}_n} \left(\frac{\tilde{L}_\pi(\omega)}{\tilde{R}_\pi(\omega)} \right) \\ &\quad \times \sup_{\omega \in \mathcal{H}_n} \left(\frac{\tilde{R}_\pi(\omega)}{R_\pi(\omega)} \right) \sup_{\omega \in \mathcal{H}_n} \left(\frac{R_\pi(\omega)}{L_{\hat{\pi}}(\omega)} \right) \frac{\tilde{L}_\pi(\hat{\omega}_{cv})}{\inf_{\omega \in \mathcal{H}_n} \tilde{L}_\pi(\omega)} - 1. \end{aligned}$$

Thus, to prove Theorem 1, it suffices to hold that

$$\sup_{\omega \in \mathcal{H}_n} \left| \frac{\tilde{R}_\pi(\omega)}{R_\pi(\omega)} - 1 \right| = o_p(1), \quad (5.12)$$

$$\sup_{\omega \in \mathcal{H}_n} \left| \frac{\tilde{L}_\pi(\omega)}{\tilde{R}_\pi(\omega)} - 1 \right| = o_p(1), \quad (5.13)$$

$$\sup_{\omega \in \mathcal{H}_n} \left| \frac{L_{\hat{\pi}}(\omega)}{R_\pi(\omega)} - 1 \right| = o_p(1), \quad (5.14)$$

$$\frac{\tilde{L}_\pi(\hat{\omega}_{cv})}{\inf_{\omega \in \mathcal{H}_n} \tilde{L}_\pi(\omega)} - 1 = o_p(1). \quad (5.15)$$

We can obtain (5.12) in Lemma 2, which implies that (5.12) is valid.

For (5.13), it is noted that

$$\begin{aligned} \left| \frac{\tilde{L}_\pi(\omega)}{\tilde{R}_\pi(\omega)} - 1 \right| &= \left| \frac{\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2 - \|\tilde{\mathbf{A}}(\omega)\boldsymbol{\mu}\|^2 - \text{tr}\{\tilde{\mathbf{P}}(\omega)\boldsymbol{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^T\}}{\|\tilde{\mathbf{P}}(\omega)\mathbf{e}_\pi\|^2 - \text{tr}\{\tilde{\mathbf{P}}(\omega)\boldsymbol{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^T\} - 2\boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega)\tilde{\mathbf{P}}(\omega)\mathbf{e}_\pi} \right| \end{aligned}$$

Hence, for (5.13) to hold, it suffices to show that

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \|\tilde{\mathbf{P}}(\omega)\mathbf{e}_\pi\|^2 - \text{tr}\{\tilde{\mathbf{P}}(\omega)\boldsymbol{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^T\} \right|}{\tilde{R}_\pi(\omega)} = o_p(1), \quad (5.16)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega)\tilde{\mathbf{P}}(\omega)\mathbf{e}_\pi \right|}{\tilde{R}_\pi(\omega)} = o_p(1). \quad (5.17)$$

In addition, according to Lemma 4 and the property of $\mathbf{P}_{(m)}$, we have

$$\lambda_{\max}\{\mathbf{P}^{(m)}\} = \lambda_{\max}\{\mathbf{Q} + \tilde{\mathbf{X}}^{(m)}(\tilde{\mathbf{X}}^{(m)T}\tilde{\mathbf{X}}^{(m)})^{-1}\tilde{\mathbf{X}}^{(m)T}\}$$

$$\begin{aligned} &\leq \lambda_{\max}\{\mathbf{Q}\} + \lambda_{\max}\{\tilde{\mathbf{X}}^{(m)}(\tilde{\mathbf{X}}^{(m)\top}\tilde{\mathbf{X}}^{(m)})^{-1}\tilde{\mathbf{X}}^{(m)\top}\} \\ &\leq 2. \end{aligned} \quad (5.18)$$

Note that

$$\tilde{\mathbf{P}}^{(m)} = \mathbf{P}^{(m)} - \mathbf{D}^{(m)}\mathbf{A}^{(m)}, \quad (5.19)$$

which, along with (5.18) and the first term of Condition (C4), leads us to

$$\begin{aligned} \lambda_{\max}\{\tilde{\mathbf{P}}(\omega)\} &\leq \sum_{m=1}^{M_n} \omega_m [\lambda_{\max}\{\mathbf{P}^{(m)}\} + \lambda_{\max}\{-\mathbf{D}^{(m)}\mathbf{A}^{(m)}\}] \\ &\leq \sum_{m=1}^{M_n} \omega_m [2 + \lambda_{\max}\{-\mathbf{D}^{(m)}\} \lambda_{\max}\{\mathbf{A}^{(m)}\}] \\ &\leq \sum_{m=1}^{M_n} \omega_m \left[2 + \max_{1 \leq i \leq n} \frac{h_{m,ii}}{1 - h_{m,ii}} \right] \\ &= 1 + (1 - \bar{h})^{-1} \\ &= O(1). \end{aligned} \quad (5.20)$$

To prove (5.16), it is necessary only to verify that, for any $\delta > 0$,

$$\begin{aligned} &P_r \left\{ \sup_{\omega \in \mathcal{H}_n} \left| \|\tilde{\mathbf{P}}(\omega)\mathbf{e}_\pi\|^2 - \text{tr}\{\tilde{\mathbf{P}}(\omega)\mathbf{\Omega}_\pi\tilde{\mathbf{P}}(\omega)^\top\} \right| / \tilde{R}_\pi(\omega) > \delta \mid X, U \right\} \\ &\leq P_r \left\{ \sup_{\omega \in \mathcal{H}_n} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} \omega_m \omega_s \left| \mathbf{e}_\pi^\top \tilde{\mathbf{P}}_{(m)}^\top \tilde{\mathbf{P}}_{(s)} \mathbf{e}_\pi - \text{tr}\{\mathbf{\Omega}_\pi \tilde{\mathbf{P}}_{(s)}^\top \tilde{\mathbf{P}}_{(m)}\} \right| > \delta \tilde{\xi}_\pi \mid X, U \right\} \\ &\leq P_r \left\{ \max_{1 \leq m \leq M_n} \max_{1 \leq s \leq M_n} \left| \mathbf{e}_\pi^\top \tilde{\mathbf{P}}_{(m)}^\top \tilde{\mathbf{P}}_{(s)} \mathbf{e}_\pi - \text{tr}\{\mathbf{\Omega}_\pi \tilde{\mathbf{P}}_{(s)}^\top \tilde{\mathbf{P}}_{(m)}\} \right| > \delta \tilde{\xi}_\pi \mid X, U \right\} \\ &\leq \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} P_r \left\{ \left| \mathbf{e}_\pi^\top \tilde{\mathbf{P}}_{(m)}^\top \tilde{\mathbf{P}}_{(s)} \mathbf{e}_\pi - \text{tr}\{\mathbf{\Omega}_\pi \tilde{\mathbf{P}}_{(s)}^\top \tilde{\mathbf{P}}_{(m)}\} \right| > \delta \tilde{\xi}_\pi \mid X, U \right\} \\ &\leq C_1 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} E \left\{ \left| \mathbf{e}_\pi^\top \mathbf{\Omega}_\pi^{-1/2} \tilde{\mathbf{P}}(\omega_m^o)^\top \tilde{\mathbf{P}}(\omega_s^o) \mathbf{\Omega}_\pi^{1/2} \mathbf{e}_\pi - \text{tr}\{\mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega_s^o)^\top \tilde{\mathbf{P}}(\omega_m^o)\} \right|^{2G} \mid X, U \right\} \\ &\leq C_1 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \lambda_{\max}^G(\mathbf{\Omega}_\pi) \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} \left| \text{tr}\{\tilde{\mathbf{P}}(\omega_m^o)^\top \tilde{\mathbf{P}}(\omega_s^o) \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega_s^o)^\top \tilde{\mathbf{P}}(\omega_m^o)\} \right|^G \\ &\leq C_1 \delta^{-2G} \lambda_{\max}^G(\mathbf{\Omega}_\pi) \lambda_{\max}^{2G}[\tilde{\mathbf{P}}(\omega_s^o)] \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \left| \text{tr}\{\tilde{\mathbf{P}}(\omega_m^o)^\top \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega_m^o)\} \right|^G \\ &\leq C_1 \delta^{-2G} \lambda_{\max}^G(\mathbf{\Omega}_\pi) \lambda_{\max}^{2G}[\tilde{\mathbf{P}}(\omega_s^o)] M_n \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \{\tilde{R}_\pi(\omega_m^o)\}^G \rightarrow 0, \text{ as } n \rightarrow \infty, \end{aligned}$$

where C_1 is a positive constant. The third inequality, fourth inequality, and fifth inequality are derived from the triangle inequality, Markov's inequality and (7) of Theorem 2 of [23], respectively. The

sixth line follows from inequality $\text{tr}(\mathbf{B}_1\mathbf{B}_2) \leq \lambda_{\max}(\mathbf{B}_1)\text{tr}(\mathbf{B}_2)$, and the last inequality is contributed to $\text{tr}\{\tilde{\mathbf{P}}(\omega_m^o)^T\mathbf{\Omega}_\pi\tilde{\mathbf{P}}(\omega_m^o)\} \leq \tilde{R}_\pi(\omega_m^o)$. The last inequality is guaranteed by Lemma 2, (5.10), and (5.20). Thus, (5.16) is valid.

By similar arguments, for (5.17), we obtain that

$$\begin{aligned} & P_r \left\{ \sup_{\omega \in \mathcal{H}_n} |\boldsymbol{\mu}^T \tilde{\mathbf{A}}^T(\omega) \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi| / \tilde{R}_\pi(\omega) > \delta \mid X, U \right\} \\ & \leq \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} P_r \left\{ |\boldsymbol{\mu}^T \tilde{\mathbf{A}}^T(\omega_m^o) \tilde{\mathbf{P}}(\omega_s^o) \mathbf{e}_\pi| > \delta \tilde{\xi}_\pi \mid X, U \right\} \\ & \leq \delta^{-2G} \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} E \left\{ |\boldsymbol{\mu}^T \tilde{\mathbf{A}}^T(\omega_m^o) \tilde{\mathbf{P}}(\omega_s^o) \mathbf{e}_\pi|^{2G} \mid X, U \right\} \\ & \leq C_2 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} |\tilde{\mathbf{P}}(\omega_s^o) \mathbf{\Omega}_\pi^{1/2} \tilde{\mathbf{A}}(\omega_m^o) \boldsymbol{\mu}|^{2G} \\ & \leq C_2 \delta^{-2G} \tilde{\xi}_\pi^{-2G} M_n \lambda_{\max}^G(\mathbf{\Omega}_\pi) \lambda_{\max}^{2G}[\tilde{\mathbf{P}}(\omega_s^o)] \sum_{m=1}^{M_n} |\tilde{\mathbf{A}}(\omega_m^o) \boldsymbol{\mu}|^{2G} \\ & \leq C_2 \delta^{-2G} \lambda_{\max}^G(\mathbf{\Omega}_\pi) \lambda_{\max}^{2G}[\tilde{\mathbf{P}}(\omega_s^o)] M_n \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \{\tilde{R}_\pi(\omega_m^o)\}^G \rightarrow 0, \text{ as } n \rightarrow \infty, \end{aligned}$$

where C_2 is a positive constant, and the last inequality is due to the fact that $\|\tilde{\mathbf{A}}(\omega_m^o) \boldsymbol{\mu}\|^2 \leq \tilde{R}_\pi(\omega_m^o)$, which is implied by (5.11). Thus, (5.17) is valid. This completes the proof of (5.13).

By the Cauchy-Schwarz inequality, it can be shown that

$$\left| \frac{L_{\hat{\pi}}(\omega)}{R_\pi(\omega)} - 1 \right| \leq \left| \frac{L_\pi(\omega)}{R_\pi(\omega)} - 1 \right| + \frac{2\{L_\pi(\omega)\}^{1/2} \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \hat{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|}{R_\pi(\omega)} + \frac{\|\hat{\boldsymbol{\mu}}_\pi(\omega) - \hat{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2}{R_\pi(\omega)}.$$

Thus, to prove (5.14), it suffices to show that

$$\sup_{\omega \in \mathcal{H}_n} \left| \frac{L_\pi(\omega)}{R_\pi(\omega)} - 1 \right| = o_p(1), \quad (5.21)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{\|\hat{\boldsymbol{\mu}}_\pi(\omega) - \hat{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2}{R_\pi(\omega)} = o_p(1). \quad (5.22)$$

Similarly, using the technique used in deriving (5.13), it can be shown that (5.21) is valid. By Lemma 1, Lemma 4, Condition (C3), and (5.18), we have

$$\begin{aligned} \sup_{\omega \in \mathcal{H}_n} \frac{\|\hat{\boldsymbol{\mu}}_\pi(\omega) - \hat{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2}{R_\pi(\omega)} &= \sup_{\omega \in \mathcal{H}_n} \frac{\|\mathbf{P}(\omega) \mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi\|^2}{R_\pi(\omega)} \\ &\leq \xi_\pi^{-1} \sup_{\omega \in \mathcal{H}_n} \lambda_{\max}^2\{\mathbf{P}(\omega)\} \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi\|^2 \\ &\leq 4\xi_\pi^{-1} \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi\|^2 \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

As a result, (5.21) and (5.22) are valid, and thus (5.14) is valid.

By the jackknife criterion in (2.2), straightforward and careful calculation yields

$$\begin{aligned} CV_{\hat{\pi}}(\omega) &= \|\mathbf{Z}_{\hat{\pi}} - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 \\ &= \|\mathbf{Z}_{\hat{\pi}} - \boldsymbol{\mu} + \boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_{\pi}(\omega) + \tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 \\ &= \tilde{L}_{\pi}(\omega) + \tilde{a}_n(\omega) + \|\mathbf{Z}_{\hat{\pi}} - \boldsymbol{\mu}\|^2, \end{aligned} \quad (5.23)$$

where the term $\|\mathbf{Z}_{\hat{\pi}} - \boldsymbol{\mu}\|^2$ is independent of ω , and

$$\begin{aligned} \tilde{a}_n(\omega) &= \|\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 + 2\{\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_{\pi}(\omega)\}^T\{\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\} + 2(\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi})^T\{\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}_{\pi}(\omega)\} \\ &\quad + 2(\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi})^T\{\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\} + 2\mathbf{e}_{\pi}^T\tilde{\mathbf{A}}(\omega)\boldsymbol{\mu} - 2\mathbf{e}_{\pi}^T\tilde{\mathbf{P}}(\omega)\mathbf{e}_{\pi} + 2\mathbf{e}_{\pi}^T\tilde{\mathbf{P}}(\omega)(\mathbf{Z}_{\pi} - \mathbf{Z}_{\hat{\pi}}). \end{aligned}$$

Thus, for (5.23) to hold, it only needs to verify that, as $n \rightarrow \infty$,

$$\sup_{\omega \in \mathcal{H}_n} \frac{|\tilde{a}_n(\omega)|}{\tilde{R}_{\pi}(\omega)} = o_p(1). \quad (5.24)$$

Using the Cauchy-Schwarz inequality, Lemma 4 and (5.20), we obtain that

$$\begin{aligned} |\tilde{a}_n(\omega)| &= \|\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 + 2\{\tilde{L}_{\pi}(\omega)\}^{\frac{1}{2}}\|\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\| + 2(\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi})^T\{\tilde{L}_{\pi}(\omega)\}^{\frac{1}{2}} \\ &\quad + 2(\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi})^T\|\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\| + 2|\mathbf{e}_{\pi}^T\tilde{\mathbf{A}}(\omega)\boldsymbol{\mu}| \\ &\quad - 2|\mathbf{e}_{\pi}^T\tilde{\mathbf{P}}(\omega)\mathbf{e}_{\pi}| + 2\|\tilde{\mathbf{P}}(\omega)^T\mathbf{e}_{\pi}\|(\mathbf{Z}_{\pi} - \mathbf{Z}_{\hat{\pi}}), \end{aligned} \quad (5.25)$$

where

$$\|\tilde{\boldsymbol{\mu}}_{\pi}(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 = \|\tilde{\mathbf{P}}(\omega)(\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi})\|^2 \leq [\lambda_{\max}\{\tilde{\mathbf{P}}(\omega)\}]^2\|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_{\pi}\|^2 = O_p(1).$$

Therefore, for (5.24) to hold, it suffices to prove that

$$\sup_{\omega \in \mathcal{H}_n} \frac{|\boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega)\mathbf{e}_{\pi}|}{\tilde{R}_{\pi}(\omega)} = o_p(1), \quad (5.26)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{|\mathbf{e}_{\pi}^T\tilde{\mathbf{P}}(\omega)\mathbf{e}_{\pi}|}{\tilde{R}_{\pi}(\omega)} = o_p(1), \quad (5.27)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{\|\tilde{\mathbf{P}}(\omega)^T\mathbf{e}_{\pi}\|}{\tilde{R}_{\pi}(\omega)} = o_p(1). \quad (5.28)$$

Likewise, the technique is applied to derive (5.17). For any $\delta > 0$, we have

$$\begin{aligned} &P_r \left\{ \sup_{\omega \in \mathcal{H}_n} |\boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega)\mathbf{e}_{\pi}| / \tilde{R}_{\pi}(\omega) > \delta \mid X, U \right\} \\ &\leq \sum_{m=1}^{M_n} P_r \left\{ |\boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega_m^o)\mathbf{e}_{\pi}| > \delta\tilde{\xi}_{\pi} \mid X, U \right\} \\ &\leq \delta^{-2G} \tilde{\xi}_{\pi}^{-2G} \sum_{m=1}^{M_n} E \left\{ |\boldsymbol{\mu}^T\tilde{\mathbf{A}}^T(\omega_m^o)\mathbf{e}_{\pi}|^{2G} \mid X, U \right\} \end{aligned}$$

$$\begin{aligned}
&\leq C_3 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \left| \Omega_\pi^{1/2} \tilde{\mathbf{A}}(\omega_m^o) \boldsymbol{\mu} \right|^{2G} \\
&\leq C_3 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \lambda_{\max}^G(\Omega_\pi) \sum_{m=1}^{M_n} \left| \tilde{\mathbf{A}}(\omega_m^o) \boldsymbol{\mu} \right|^{2G} \\
&\leq C_3 \delta^{-2G} \lambda_{\max}^G(\Omega_\pi) \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \{\tilde{R}_\pi(\omega_m^o)\}^G \rightarrow 0, \text{ as } n \rightarrow \infty,
\end{aligned}$$

where C_3 is a positive constant. By Conditions (C2) and (C3) and (5.10), we know that (5.26) is valid.

It can be observed that

$$\left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi \right| \leq \left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi - \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi) \right| + \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi).$$

Therefore, (5.27) holds if we can prove that

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi - \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi) \right|}{\tilde{R}_\pi(\omega)} = o_p(1), \quad (5.29)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi) \right|}{\tilde{R}_\pi(\omega)} = o_p(1). \quad (5.30)$$

Similar to (5.26), it can be shown that

$$\begin{aligned}
&P_r \left\{ \sup_{\omega \in \mathcal{H}_n} \left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi - \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi) \right| / \tilde{R}_\pi(\omega) > \delta \mid X, U \right\} \\
&\leq \sum_{m=1}^{M_n} P_r \left\{ \left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega_m^o) \mathbf{e}_\pi - \text{tr}(\tilde{\mathbf{P}}(\omega_m^o) \Omega_\pi) \right| > \delta \tilde{\xi}_\pi \mid X, U \right\} \\
&\leq \delta^{-2G} \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} E \left\{ \left| \mathbf{e}_\pi^T \Omega_\pi^{-1/2} \Omega_\pi^{1/2} \tilde{\mathbf{P}}(\omega_m^o) \Omega_\pi^{1/2} \Omega_\pi^{-1/2} \mathbf{e}_\pi - \text{tr}(\Omega_\pi \tilde{\mathbf{P}}(\omega_m^o)) \right|^{2G} \mid X, U \right\} \\
&\leq C_4 \delta^{-2G} \tilde{\xi}_\pi^{-2G} \lambda_{\max}^G(\Omega_\pi) \sum_{m=1}^{M_n} \left(\text{tr} \left\{ \tilde{\mathbf{P}}(\omega_m^o)^T \Omega_\pi \tilde{\mathbf{P}}(\omega_m^o) \right\} \right)^G \\
&\leq C_4 \delta^{-2G} \lambda_{\max}^G(\Omega_\pi) \tilde{\xi}_\pi^{-2G} \sum_{m=1}^{M_n} \{\tilde{R}_\pi(\omega_m^o)\}^G \rightarrow 0, \text{ as } n \rightarrow \infty,
\end{aligned}$$

where C_4 is a positive constant. As a result, (5.29) is valid.

By (5.10), Condition (C4), and the fact that all the diagonal elements of $\tilde{\mathbf{P}}^{(m)}$ are zeros, it is observed that

$$\begin{aligned}
\sup_{\omega \in \mathcal{H}_n} \left| \text{tr}(\tilde{\mathbf{P}}(\omega) \Omega_\pi) \right| / \tilde{R}_\pi(\omega) &\leq \tilde{\xi}_\pi^{-1} \max_{1 \leq m \leq M_n} \left| \text{tr}(\tilde{\mathbf{P}}^{(m)} \Omega_\pi) \right| \\
&\leq \tilde{\xi}_\pi^{-1} \max_{1 \leq m \leq M_n} \left| \lambda_{\max}(\Omega_\pi) \text{tr}(\tilde{\mathbf{P}}^{(m)}) \right|
\end{aligned}$$

$$\leq \tilde{\xi}_\pi^{-1} \lambda_{\max}(\mathbf{\Omega}_\pi) \max_{1 \leq m \leq M_n} \text{tr}(\tilde{\mathbf{P}}^{(m)}) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Thus, (5.30) is valid.

Similarly, for (5.28), we have that

$$\|\tilde{\mathbf{P}}(\omega)^T \mathbf{e}_\pi\|^2 \leq \left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega)^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi - \text{tr} \{ \tilde{\mathbf{P}}(\omega)^T \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega) \} \right| + \text{tr} \{ \tilde{\mathbf{P}}(\omega)^T \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega) \}.$$

Then, for (5.28) to hold, we only need to prove

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \mathbf{e}_\pi^T \tilde{\mathbf{P}}(\omega)^T \tilde{\mathbf{P}}(\omega) \mathbf{e}_\pi - \text{tr} \{ \tilde{\mathbf{P}}(\omega)^T \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega) \} \right|}{\{\tilde{R}_\pi(\omega)\}^2} = o_p(1), \quad (5.31)$$

$$\sup_{\omega \in \mathcal{H}_n} \frac{\left| \text{tr} \{ \tilde{\mathbf{P}}(\omega)^T \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega) \} \right|}{\{\tilde{R}_\pi(\omega)\}^2} = o_p(1). \quad (5.32)$$

By the proof of (5.16), it can be shown that (5.31) is valid. Letting $\mathbf{S}^{(m)} = \mathbf{D}^{(m)} + \mathbf{I}_n$, this together with (5.19) generates

$$\begin{aligned} \text{tr} \{ \tilde{\mathbf{P}}^{(m)T} \tilde{\mathbf{P}}^{(m)} \} &= \text{tr} \{ [\mathbf{P}^{(m)} - \mathbf{D}^{(m)} \mathbf{A}^{(m)}] \tilde{\mathbf{P}}^{(m)} \} \\ &= \text{tr} \{ [(\mathbf{P}^{(m)} - \mathbf{I}_n) \mathbf{S}^{(m)} + \mathbf{I}_n] \tilde{\mathbf{P}}^{(m)} \} \\ &= \text{tr} \{ \mathbf{P}^{(m)} \mathbf{S}^{(m)} \tilde{\mathbf{P}}^{(m)} \} - \text{tr} \{ \mathbf{S}^{(m)} \tilde{\mathbf{P}}^{(m)} \} \\ &= \text{tr} \{ \mathbf{P}^{(m)} \mathbf{S}^{(m)} \mathbf{S}^{(m)} (\mathbf{P}^{(m)} - \mathbf{I}_n) \} + \text{tr} \{ \mathbf{P}^{(m)} \mathbf{S}^{(m)} \} \\ &\leq \text{tr} \{ \mathbf{P}^{(m)} \mathbf{S}^{(m)} \mathbf{S}^{(m)} \mathbf{P}^{(m)} \} + \text{tr} \{ \mathbf{P}^{(m)} \mathbf{S}^{(m)} \} \\ &\leq \text{tr} \{ \mathbf{P}^{(m)} (1 - \bar{h})^{-2} \} + \text{tr} \{ \mathbf{P}^{(m)} (1 - \bar{h})^{-1} \} \\ &= (d_n + k_m) (1 - \bar{h})^{-2} (2 - \bar{h}), \end{aligned} \quad (5.33)$$

where

$$\begin{aligned} \text{tr} \{ \mathbf{P}^{(m)} \} &= \text{tr} \{ \mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \} \\ &= \text{tr} \{ \mathbf{Q} \} + \text{tr} \{ \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \} \\ &= d_n + k_m, \\ \text{tr} \{ \tilde{\mathbf{P}}^{(m)} \} &= 0. \end{aligned}$$

Under Lemma 2 and the second part of Condition (C4), we have

$$\tilde{\xi}_\pi^{-2} (\bar{d} + \bar{k}) = \xi_\pi^{-2} (\bar{d} + \bar{k}) \xi_\pi^2 \tilde{\xi}_\pi^{-2} \leq \xi_\pi^{-2} (\bar{d} + \bar{k}) \left\{ \sup_{\omega \in \mathcal{H}_n} \left| \frac{R_\pi(\omega)}{\tilde{R}_\pi(\omega)} - 1 \right| + 1 \right\} \xrightarrow{a.s.} 0.$$

where $\tilde{\xi}_\pi = \inf_{\omega \in \mathcal{H}_n} \tilde{R}(\omega)$. Which, along with (5.10) and (5.33), implies that

$$\sup_{\omega \in \mathcal{H}_n} \left| \text{tr} \{ \tilde{\mathbf{P}}(\omega)^T \mathbf{\Omega}_\pi \tilde{\mathbf{P}}(\omega) \} \right| / \{\tilde{R}_\pi(\omega)\}^2$$

$$\begin{aligned} &\leq \tilde{\xi}_\pi^{-2} \sup_{\omega \in \mathcal{H}_n} \left| \lambda_{\max}(\mathbf{\Omega}_\pi) \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} \omega_m \omega_s \operatorname{tr} \left\{ \tilde{\mathbf{P}}^{(m)\top} \tilde{\mathbf{P}}^{(s)} \right\} \right| \\ &\leq \lambda_{\max}(\mathbf{\Omega}_\pi) \tilde{\xi}_\pi^{-2} (\bar{d} + \bar{k})(1 - \bar{h})^{-2} (2 - \bar{h}) \rightarrow 0, \text{ as } n \rightarrow \infty, \end{aligned}$$

and thus (5.32) is valid. In conclusion, the proof of Theorem 1 is completed. \square

Proof of Theorem 2:

Define $\boldsymbol{\psi}_n(\boldsymbol{\omega}) = \mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi + \hat{\boldsymbol{\mu}}_\pi(\boldsymbol{\omega}) - \tilde{\boldsymbol{\mu}}_\pi(\boldsymbol{\omega}) + \tilde{\boldsymbol{\mu}}_\pi(\boldsymbol{\omega}) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\boldsymbol{\omega})$. A simple calculation yields

$$\begin{aligned} CV_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_{cv}) &= \|\mathbf{Z}_{\hat{\pi}} - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_{cv})\|^2 \\ &= \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi + \mathbf{Z}_\pi - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_{cv}) + \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_{cv}) - \tilde{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_{cv}) + \tilde{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_{cv}) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_{cv})\|^2 \\ &= \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_{cv}) + \mathbf{e}_\pi + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv})\|^2 \\ &= \left\| \hat{\delta}_{cv} \sum_{m=1}^{m_0} \frac{\hat{\boldsymbol{\omega}}_{cv,m}}{\hat{\delta}_{cv}} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi^{(m)}\} + (1 - \hat{\delta}_{cv}) \sum_{m=m_0+1}^{M_n} \frac{\hat{\boldsymbol{\omega}}_{cv,m}}{1 - \hat{\delta}_{cv}} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi^{(m)}\} + \mathbf{e}_\pi + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv}) \right\|^2 \\ &= \left\| \hat{\delta}_{cv} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\hat{\boldsymbol{\omega}}_C)\} + (1 - \hat{\delta}_{cv}) \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(\hat{\boldsymbol{\omega}}_F)\} + \mathbf{e}_\pi + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv}) \right\|^2, \end{aligned} \quad (5.34)$$

where $\hat{\boldsymbol{\omega}}_C = (\hat{\boldsymbol{\omega}}_{cv,1}, \dots, \hat{\boldsymbol{\omega}}_{cv,m_0}, 0, \dots, 0) / \hat{\delta}_{cv} \in \mathcal{H}_n$ and $\hat{\boldsymbol{\omega}}_F = (0, \dots, 0, \hat{\boldsymbol{\omega}}_{cv,m_0+1}, \dots, \hat{\boldsymbol{\omega}}_{cv,M_n}) / (1 - \hat{\delta}_{cv}) \in \mathcal{H}_n$. Likewise, we obtain that

$$\begin{aligned} CV_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_C) &= \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi + \mathbf{Z}_\pi - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C) + \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C) - \tilde{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C) + \tilde{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_C)\|^2 \\ &= \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C) + \mathbf{e}_\pi + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_C)\|^2. \end{aligned} \quad (5.35)$$

We know that $CV_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_{cv}) \leq CV_{\hat{\pi}}(\hat{\boldsymbol{\omega}}_C)$, which with (5.34), (5.35), and the Cauchy-Schwarz inequality, implies that

$$\begin{aligned} (1 - \hat{\delta}_{cv})^2 &\leq \left[(1 - \hat{\delta}_{cv}^2) \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\|^2 + \|\boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_C)\|^2 + 2\mathbf{e}_\pi^\top [\{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\} + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_C)] \right. \\ &\quad + 2\{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\}^\top \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_C) + 2[\hat{\delta}_{cv} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\} + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv})]^\top \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_F)\} \\ &\quad + \|\boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv})\|^2 + 2\mathbf{e}_\pi^\top \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_F)\} + 2\mathbf{e}_\pi^\top [\hat{\delta}_{cv} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\} + \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv})] \\ &\quad \left. + 2\hat{\delta}_{cv} \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\}^\top \boldsymbol{\psi}_n(\hat{\boldsymbol{\omega}}_{cv}) \right] / \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_F)\| \\ &\leq \left[2\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\|^2 + 2 \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\|^2 + 4 \left| \mathbf{e}_\pi^\top \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\} \right| + 4\|\mathbf{e}_\pi\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right. \\ &\quad + 4\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| + 2 \left\{ \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\| + \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right\} \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_F)\| \\ &\quad \left. + 2\mathbf{e}_\pi^\top \mathbf{A}(\hat{\boldsymbol{\omega}}_F) \boldsymbol{\mu} - 2\mathbf{e}_\pi^\top \mathbf{P}(\hat{\boldsymbol{\omega}}_F) \mathbf{e}_\pi \right] \frac{1}{R_\pi(\hat{\boldsymbol{\omega}}_F)} \frac{R_\pi(\hat{\boldsymbol{\omega}}_F)}{L_\pi(\hat{\boldsymbol{\omega}}_F)} \\ &\leq \left[\left\{ 2\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\|^2 + 2 \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\|^2 + 4 \left| \mathbf{e}_\pi^\top \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\} \right| + 4\|\mathbf{e}_\pi\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right. \right. \\ &\quad \left. \left. + 4\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\boldsymbol{\omega}}_C)\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right\} \xi_{\pi,F}^{-1} + 2 \frac{\mathbf{e}_\pi^\top \mathbf{A}(\hat{\boldsymbol{\omega}}_F) \boldsymbol{\mu}}{R_\pi(\hat{\boldsymbol{\omega}}_F)} + 2 \frac{\mathbf{e}_\pi^\top \mathbf{P}(\hat{\boldsymbol{\omega}}_F) \mathbf{e}_\pi}{R_\pi(\hat{\boldsymbol{\omega}}_F)} \right] \end{aligned}$$

$$\begin{aligned}
& + 2 \left\{ \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\| + \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right\} \left\{ \frac{L_{\pi}(\hat{\omega}_F)}{R_{\pi}(\hat{\omega}_F)} \right\}^{1/2} \xi_{\pi,F}^{-1/2} \sup_{\omega \in \mathcal{H}_F} \left[\left| \frac{L_{\pi}(\omega)}{R_{\pi}(\omega)} - 1 \right| + 1 \right] \\
\leq & \left\{ 2\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\|^2 + 2 \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\|^2 + 4 \left| \mathbf{e}_{\pi}^T \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\} \right| + 4 \|\mathbf{e}_{\pi}\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right. \\
& + 4 \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\| \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \left. \right\} \xi_{\pi,F}^{-1} + 2 \sup_{\omega \in \mathcal{H}_F} \frac{|\mathbf{e}_{\pi}^T \mathbf{A}(\omega) \boldsymbol{\mu}|}{R_{\pi}(\omega)} + 2 \sup_{\omega \in \mathcal{H}_F} \frac{\mathbf{e}_{\pi}^T \mathbf{P}(\omega) \mathbf{e}_{\pi}}{R_{\pi}(\omega)} \\
& + 2 \xi_{\pi,F}^{-1/2} \left\{ \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\| + \sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\| \right\} \left\{ \sup_{\omega \in \mathcal{H}_F} \left| \frac{L_{\pi}(\omega)}{R_{\pi}(\omega)} - 1 \right| + 1 \right\}^{1/2} \\
& \times \sup_{\omega \in \mathcal{H}_F} \left\{ \left| \frac{L_{\pi}(\omega)}{R_{\pi}(\omega)} - 1 \right| + 1 \right\}.
\end{aligned}$$

Condition (C6) indicates that to prove Theorem 2, it suffices to show

$$\xi_{\pi,F}^{-1} \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\|^2 = o_p(1), \quad (5.36)$$

$$\xi_{\pi,F}^{-1} \left| \mathbf{e}_{\pi}^T \{\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\} \right| = o_p(1), \quad (5.37)$$

$$\sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\|^2 = O_p(1), \quad (5.38)$$

$$\xi_{\pi,F}^{-2} \|\mathbf{e}_{\pi}\|^2 = o_p(1), \quad (5.39)$$

$$\sup_{\omega \in \mathcal{H}_F} \left| \frac{L_{\pi}(\omega)}{R_{\pi}(\omega)} - 1 \right| = o_p(1), \quad (5.40)$$

$$\sup_{\omega \in \mathcal{H}_F} \frac{|\mathbf{e}_{\pi}^T \mathbf{A}(\omega) \boldsymbol{\mu}|}{R_{\pi}(\omega)} = o_p(1), \quad (5.41)$$

$$\sup_{\omega \in \mathcal{H}_F} \frac{\mathbf{e}_{\pi}^T \mathbf{P}(\omega) \mathbf{e}_{\pi}}{R_{\pi}(\omega)} = o_p(1). \quad (5.42)$$

For the correct model with $m = 1, 2, \dots, m_0$, we have

$$\begin{aligned}
\mathbf{P}^{(m)} \boldsymbol{\mu} &= \left\{ \mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \right\} \left\{ \mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \right\} \mathbf{Z}_{\pi} \\
&= \left\{ \mathbf{Q} + 2\mathbf{Q}\tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \right. \\
&\quad \left. + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \right\} \mathbf{Z}_{\pi} \\
&= \left\{ \mathbf{Q} + \tilde{\mathbf{X}}^{(m)} (\tilde{\mathbf{X}}^{(m)T} \tilde{\mathbf{X}}^{(m)})^{-1} \tilde{\mathbf{X}}^{(m)T} \right\} \mathbf{Z}_{\pi} \\
&= \boldsymbol{\mu},
\end{aligned}$$

where $\mathbf{Q}\tilde{\mathbf{X}}^{(m)} = \mathbf{Q}(\mathbf{I} - \mathbf{Q})\tilde{\mathbf{X}}^{(m)} = \mathbf{0}$. This implies that

$$\begin{aligned}
\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_{\pi}(\hat{\omega}_C)\|^2 &= \left\| \sum_{m=1}^{m_0} \frac{\hat{\omega}_{cv,m}}{\hat{\delta}_{cv}} \{\boldsymbol{\mu} - \mathbf{P}^{(m)} \mathbf{Z}_{\pi}\} \right\|^2 \\
&= \left\| \sum_{m=1}^{m_0} \frac{\hat{\omega}_{cv,m}}{\hat{\delta}_{cv}} \mathbf{P}^{(m)} \mathbf{e}_{\pi} \right\|^2
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{2} \sum_{m=1}^{m_0} \frac{\hat{\omega}_{cv,m}}{\hat{\delta}_{cv}} \sum_{t=1}^{m_0} \frac{\hat{\omega}_{cv,s}}{\hat{\delta}_{cv}} (\mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{P}^{(m)} \mathbf{e}_\pi + \mathbf{e}_\pi^T \mathbf{P}_{(s)} \mathbf{P}_{(s)} \mathbf{e}_\pi) \\
&= \sum_{m=1}^{m_0} \frac{\hat{\omega}_{cv,m}}{\hat{\delta}_{cv}} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{P}^{(m)} \mathbf{e}_\pi \\
&\leq \max_{1 \leq m \leq m_0} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi.
\end{aligned}$$

Thus, for (5.36) to hold, we need to make

$$\xi_{\pi,F}^{-1} \max_{1 \leq m \leq m_0} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi = o_p(1). \quad (5.43)$$

By Markov's inequality, for any $\delta > 0$, we have

$$\begin{aligned}
\sup_{n \geq 1} P \left(\max_{1 \leq m \leq m_0} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi > \delta \right) &\leq \sup_{n \geq 1} \sum_{m=1}^{m_0} P \left(\mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi > \delta \right) \\
&= \sup_{n \geq 1} \sum_{m=1}^{m_0} E \left[E \left\{ I \left(\mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi > \delta \right) \middle| X, U \right\} \right] \\
&= \sup_{n \geq 1} \sum_{m=1}^{m_0} E \left\{ P \left(\mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi > \delta \middle| X, U \right) \right\} \\
&\leq \sup_{n \geq 1} \sum_{m=1}^{m_0} E \left[\delta^{-2G} E \left\{ \left(\mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi \right)^{2G} \middle| X, U \right\} \right] \\
&\leq \sup_{n \geq 1} \sum_{m=1}^{m_0} E \left\{ \delta^{-2G} \text{tr} \left\{ \mathbf{P}^{(m)} \mathbf{\Omega}_\pi \right\}^{2G} \right\} \\
&\leq \delta^{-2G} \lambda_{\max}^{2G}(\mathbf{\Omega}) m_0 (\bar{d} + \bar{k})^{2G},
\end{aligned}$$

which is $o_p(1)$ under Condition (C6), implying that (5.43) is valid and thus guarantees that (5.36) is valid.

Indeed, (5.37) is further simplified to the following form:

$$\left| \mathbf{e}_\pi^T \{ \boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_\pi(\hat{\omega}_C) \} \right| = \left| \sum_{m=1}^{m_0} \frac{\hat{\omega}_{cv,m}}{\hat{\delta}_{cv}} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi \right| \leq \max_{1 \leq m \leq m_0} \mathbf{e}_\pi^T \mathbf{P}^{(m)} \mathbf{e}_\pi.$$

Thus, by the proof of (5.43), it can be shown that (5.37) is valid.

For (5.38), one can obtain that

$$\begin{aligned}
\sup_{\omega \in \mathcal{H}_n} \|\boldsymbol{\psi}_n(\omega)\|^2 &= \sup_{\omega \in \mathcal{H}_n} \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi + \hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega) + \tilde{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 \\
&\leq 2 \sup_{\omega \in \mathcal{H}_n} \left\{ \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi\|^2 + \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2 + \|\tilde{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\omega)\|^2 \right\} \\
&\leq 2 \{ 1 + (1 - \bar{h})^{-2} \} \|\mathbf{Z}_{\hat{\pi}} - \mathbf{Z}_\pi\|^2 + 2 \sup_{\omega \in \mathcal{H}_n} \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2 \\
&= O_p(1) + 2 \sup_{\omega \in \mathcal{H}_n} \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2, \quad (5.44)
\end{aligned}$$

where the third step is derived from the Cauchy-Schwarz inequality and (5.25), and the last step is obtained from Lemma 2 and Condition (C7). Therefore, (5.44) holds if we can prove that

$$\sup_{\omega \in \mathcal{H}_n} \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2 = o_p(1). \quad (5.45)$$

By (5.19), Lemma 4, and the Cauchy-Shwarz inequality, we have

$$\begin{aligned} \sup_{\omega \in \mathcal{H}_n} \|\hat{\boldsymbol{\mu}}_\pi(\omega) - \tilde{\boldsymbol{\mu}}_\pi(\omega)\|^2 &= \sup_{\omega \in \mathcal{H}_n} \|\mathbf{P}(\omega)\mathbf{Z}_\pi - \tilde{\mathbf{P}}(\omega)\mathbf{Z}_\pi\|^2 \\ &= \sup_{\omega \in \mathcal{H}_n} \left\| \sum_{m=1}^{M_n} \omega_m \mathbf{D}^{(m)} \mathbf{A}^{(m)} \mathbf{Z}_\pi \right\|^2 \\ &\leq \frac{1}{2} \sup_{\omega \in \mathcal{H}_n} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} \omega_m \omega_s \left(\mathbf{Z}_\pi^\top \mathbf{A}^{(m)} \mathbf{D}^{(m)} \mathbf{D}^{(s)} \mathbf{A}^{(s)} \mathbf{Z}_\pi \right. \\ &\quad \left. + \mathbf{Z}_\pi^\top \mathbf{A}^{(s)} \mathbf{D}^{(s)} \mathbf{D}^{(m)} \mathbf{A}^{(m)} \mathbf{Z}_\pi \right) \end{aligned} \quad (5.46)$$

$$\begin{aligned} &\leq \sup_{\omega \in \mathcal{H}_n} \sum_{m=1}^{M_n} \sum_{s=1}^{M_n} \omega_m \omega_s \lambda_{\max}\{\mathbf{A}^{(m)} \mathbf{D}^{(m)} \mathbf{D}^{(s)} \mathbf{A}^{(s)}\} \mathbf{Z}_\pi^\top \mathbf{Z}_\pi \\ &\leq 2(1 - \bar{h})^{-2} \bar{h}^2 n \left(\frac{1}{n} \boldsymbol{\mu}^\top \boldsymbol{\mu} + \frac{1}{n} \mathbf{e}_\pi^\top \mathbf{e}_\pi \right), \end{aligned} \quad (5.47)$$

where $\lambda_{\max}(\mathbf{A}) = -1$. This, along with Condition (C2), Condition (C4), and (5.46), implies that, to prove (5.44), it only suffices to prove

$$\frac{1}{n} \mathbf{e}_\pi^\top \mathbf{e}_\pi = O_p(1). \quad (5.48)$$

Likewise, we use the same technique used in Lemma 1 in deriving (5.48). Thus, it is valid and obtaining (5.38) is valid. Based on (5.48) and Condition (3), we know that (5.39) is valid.

Under Conditions (C6) and (C7), it is observed that

$$(M_n - m_0) \xi_{\pi, F}^{-2G} \sum_{m=m_0}^{M_n - m_0} \{R_\pi(\omega_m^o)\}^G \rightarrow 0, \quad \bar{h} \rightarrow 0, \quad (\bar{d} + \bar{k}) \xi_{F, \pi}^{-2} \rightarrow 0, \quad \text{a.s.}$$

This result implies that Conditions (C3) and (C4) are satisfied for $\{\mathbf{Z}_{\pi, i}, X_i, U_i\}_{i=1}^n$ with $\omega \in \mathcal{H}_F$. Therefore, it is analogous to (5.13), and we directly obtain that (5.40)–(5.42) are valid. Above all, this completes the proof of Theorem 2. \square



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)