Estimation and Inference of Semiparametric Models using Data from Several Sources

Moshe Buchinsky†   Fanghua Li†   Zhipeng Liao§

This Version: September 2018

Abstract

This paper studies the estimation and inference of nonlinear econometric models when the economic variables are contained in different data sets. We construct a semiparametric minimum distance (SMD) estimator of the unknown structural parameter of interest when there are some common conditioning variables in different data sets. The SMD estimator is shown to be consistent and has an asymptotic normal distribution. We provide the specific form of optimal weight for the SMD estimation, and show that the optimal weighted SMD estimator has the smallest asymptotic variance among all SMD estimators. A consistent estimator of the variance-covariance matrix of the SMD estimator, and hence inference procedure of the unknown parameter are also provided. The finite sample performances of the SMD estimators and the inference procedure are investigated in simulation studies.

Keywords: Conditional Moment Restrictions; Data Combination; Minimum Distance Estimation, Series Estimation

1 Introduction

There are many cases in empirical micro studies where data that is needed to analyze a particular phenomenon is not always available in one data set. Typically, this hampers the possibility of meaningful empirical research. In fact, a common practice is to make some simplifying assumptions, which would then permit the researcher to use information from more than one data source. For

*We wish to thank Gary Chamberlain, Xiaohong Chen, Yuichi Kitamura, Steve Berry, Richard Blundell, Don Andrews, Whitney Newey, Denis Chetverikov, Iliana Reggio, Shuyang Sheng, Rosa Matzkin, Bob Miller, Joshua Angrist, and Chuck Manski, for many good suggestions and comprehensive discussions. We also also wish to thank seminar participants at UCLA, USC, Barcelona GSE Summer Form, and Harvard University for many good comments and suggestions.

†Department of Economics, UC Los Angeles, 8373 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: buchinsky@econ.ucla.edu

‡Department of Economics, University of New South Weal4s, Australia.

§Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu
example, as Blundell et al. (2008) note, this is a crucial difficulty when faced by those studying households’ consumption and saving behavior because of the lack of panel data on household expenditures, income, and saving in one data set.

Important data for studying consumption is the Panel Study of Income Dynamics (PSID), a survey that provides longitudinal annual data for households that have been followed since 1968. The PSID collects data on a subset of consumption items, namely food eaten at home, food eaten away from home (with a few gaps in some of the survey years) and income. However, the problem with the PSID is that it does not provide data on wealth. In contrast, there are a few data sets that provide detailed data on income and wealth (e.g. Health and Retirement Survey (HRS), or the National Longitudinal Study (NLS)), but these data sets provide no information on consumption.

Problems of similar nature exist in many other countries. For example, in the UK, the Family Expenditure Survey (FES) provides comprehensive data on household expenditures, but this is across-sectional data and thus the researcher does not get to observe households over time. In contrast, the British Household Panel Survey (BHPS) is a Panel data set that collects data on income or wealth, but it provides no information on consumption. This is quite puzzling, given the vital need to study the consumption decisions jointly with the income and wealth processes.

Consequently, as is comprehensively explained in Blundell et al. (2008), studies aimed at understanding consumption behavior, and testing alternative theories, have resorted to the limited data on food expenditure provided in the PSID. These studies include, for example: Hall and Mishkin (1982), Zeldes (1989), Runkle (1991), Shea (1995), Cochrane (1991), Hayashi et al. (1996), Hall and Mishkin (1982), Waldkirch et al. (2004), Martin (2003) and Hurst and Stafford (2004). The main problem with all of these studies is that they use consumption on limited number of goods (largely necessity goods), thus putting into question the external validity of the results.

One solution in the literature is to form synthetic panel data sets from repeated cross-section data sets in which consumption is reported (e.g. the CEX or the FES). This is done in, for example, Browning et al. (1985) and Attanasio and Weber (1993). An alternative empirical approach that has been used occasionally in the literature involves imputation of consumption to the PSID households using information on consumption from the CEX. Specifically, Skinner (1987) proposes to impute total consumption in the PSID using the estimated coefficients of a regression of total consumption on a number of consumption items that are reported in both the PSID and the CEX. While this method seems appealing at first sight, it reduces any variation in total consumption, since it does not take into account the fact the there is considerable idiosyncratic elements that goes into an individual’s decision making. Ziliak (1998) and Browning and Leth-Petersen (2003) provide an alternative method that is a variant of that proposed by Skinner (1987). However, this method has some a major weakness, in that it ignored, by construction, the dynamics of the individuals’

---

1 Similar problems exist for many other countries, in particular countries in Europe (e.g. France and Spain) that collect detailed data on consumption, income, and wealth, but the information never exists in a single data set.
consumption. Avoiding direct control of an individual’s heterogeneity has been shown to create a major obstacle when modeling an individual’s behavior in general.

One paper in the literature that provides a method similar in spirit to the method proposed here is Blundell et al. (2008). The method is somewhat similar to the one proposed by Skinner (1987), in that the authors impute consumption data for the households in the PSID using regression parameters estimated from the CEX data. The key difference, is in Blundell et al. (2008) they use “structural” regression of a standard demand function for food that depends not only on other consumption items, but also depends on prices and a set of demographic and socio-economic variables of the household. Assuming monotonicity of the demand function makes it possible then to invert this function in order to obtain a structurally based formula for non-durable consumption (that exists in the CEX, but is missing in the PSID). This imputation method has some problems. First, the consumption data imputed in this fashion is likely to suffer from the well-known error-in-variable problem. Second, the method does not allow for one to account for the inherent heterogeneity in individuals’ consumption.

Two recent papers, Fan et al. (2014) and Fan et al. (2016) tackle a special case of the problem addressed in our paper, namely the case of treatment effect. Under this scenario, the outcome variables and conditioning variables are observed in two separate data sets, so that the treatment effect parameters are not point-identified. The authors provide sharp bounds on the counterfactual distributions and parameters of interest (see Fan et al. (2016)), and the corresponding inference (see Fan et al. (2014)).

Our case encompasses a more general situation in which some of the variables are available only in one data set, while others are available in a separate data sets. The key insight is that there are some variables that appear in both. Under relatively mild regularity conditions we provide a method that allows one to point-identify the structural parameters of interest in the main data set of interest using the information provided in the other data sets. The parameters of interest and the “imputation equation” are estimated simultaneously. We also provide the necessary theory for inference including cases in which the number of observation in the “imputation” data set do not diverge to infinity at the same rate as that for the main data set of interest.

There is a large literature on estimating econometric models with different data sets. The semiparametric estimator proposed in this paper is related to the two sample instrumental variable (2SIV) estimator studied in Klevenmarken (1982), Angrist and Krueger (1992) and Arellano and Meghir (1992) and the two-sample GMM estimator studied in Ridder and Mollitt (2007). The main difference is that these papers consider the estimation of the structural parameters where there are a finite number of moment restrictions, while our paper studies the estimation problem using conditional moment conditions, and thus allow for infinitely many unconditional moment restrictions. The other related works include Hellerstein and Imbens (1999), Chen et al. (2005) and

---

This method is used extensively in Blundell, Pistaferri and Preston (2008).
Chen et al. (2008). We refer the readers to Ridder and Moffitt (2007) for a recent survey on this topic.

The rest of the paper is organized as follows. Section 2 discusses the model and provides the semiparametric minimum distance (SMD) estimator. Section 3 presents the asymptotic properties of the proposed estimator. Section 4 studies the optimal weighting in the two sample minimum distance estimation. Consistent estimators of the optimal weight and the variance of the SMD are provided in Section 5. Section 6 investigates the finite sample performance of the SMD estimator and related inference procedures using Monte Carlo simulations. Section 7 concludes. Proof of the main results of the paper is given in the Appendix.

The notation used in this paper is standard. Throughout this paper, we use \( C \) to denote a generic positive finite constant which is larger than 1. For any real symmetric matrix \( A \), \( \lambda_{\text{min}}(A) \) and \( \lambda_{\text{max}}(A) \) denote the smallest and largest eigenvalues of \( A \) respectively. \( tr(\cdot) \) denotes the trace operator of square matrices. For any matrix \( A \), \( A' \) refers to the transpose of \( A \). For any real matrices \( A_1 \) and \( A_2 \), \( A_1 \otimes A_2 \) denotes the Kronecker product of \( A_1 \) and \( A_2 \). The notations \( \| \cdot \| \) and \( \| \cdot \|_{op} \) denote the Frobenius norm and the operator norm of matrices. For any set of real vectors \( \{a_i\}_{i \in I} \) with the same dimension where \( I = \{i_1, \ldots, i_d\} \) is an index set with \( d_I \) distinct natural numbers, we define \((a_i)_{i \in I} = (a_{i_1}, \ldots, a_{i_d})\) and \((a_i)'_{i \in I} = (a_{i_1}, \ldots, a_{i_d})'\). \( I_k, 0_k \) and \( 0_{k \times k} \) are used to denote \( k \times 1 \) zero vector and \( k \times k \) zero matrices respectively. The symbolism \( a \equiv b \) means that \( a \) is defined as \( b \). We use \( \mathbb{N} \) to denote the set of natural numbers. For any (possibly random) positive sequences \( \{a_n\}_{n=1}^\infty \) and \( \{b_n\}_{n=1}^\infty \), \( a_n = O_p(b_n) \) means that \( \sup_{n \in \mathbb{N}} \Pr(a_n/b_n > M) \to 0 \) as \( M \to \infty \); \( a_n = o_p(b_n) \) means that for all \( \varepsilon > 0 \), \( \lim_{n \to \infty} \Pr(a_n/b_n > \varepsilon) = 0 \). As usual, "\( \to \)" and "\( \to_d \)" imply convergence in probability and convergence in distribution, respectively.

2 The Model and the Estimator

We are interested in estimating a structural parameter \( \theta_0 \) identified by the following conditional moment restrictions

\[
E[g_1(Y_1, \theta_0) - g_2(Y_2, \theta_0)|Z] = 0_{d_y}
\]  

(1)

where \( Y_j \) \((j = 1, 2)\) are \( d_j \times 1 \) random vectors which may include both endogenous and exogenous variables, \( Z \) is a vector of instrumental variables (IVs) with support \( Z \), and \( g_j(\cdot, \cdot) : R^{d_j} \times R^{d_y} \to R^{d_y} \) \((j = 1, 2)\) are known functions. Two data sets are available to estimate the unknown parameter \( \theta_0 \): \( \{(Y'_{1,i}, Z'_i)\}_{i \in I_1} \) and \( \{(Y'_{2,i}, Z'_i)\}_{i \in I_2} \), where \( I_1 \) and \( I_2 \) are two index sets with cardinalities \( n_1 \) and \( n_2 \) respectively. Throughout the paper, we use \( I_j \) \((j = 1, 2)\) to denote both the data set \( j \) and its index set.

The unknown parameter \( \theta_0 \) could be conveniently estimated under the conditional moment restriction (1), if we had joint observations on \( (Y', Z') \), where \( Y' = (Y'_1, Y'_2) \). Many methods in the
literature can be used in such cases, for example the generalized method of moment estimation (see, e.g., [Donald et al. (2003)], the empirical likelihood estimation (see, e.g., [Donald et al. (2003) and Kitamura and Ahn (2004)]) and the minimum distance estimation (see, e.g., [Ai and Chen (2003), Chen and Pouzo (2009), Chen and Pouzo (2012) and Chen and Pouzo (2015)]). However, such straightforward methods are not applicable here because the components in $Y$ are contained in different data sets. On the other hand, since the IVs $Z$ are contained in both data sets, they can be employed for identifying and estimating the unknown parameter $\theta_0$. One way of using the common variables $Z$ in both data sets is through the data imputation. That is, one runs a (parametric or nonparametric) regression of $Y_{j1}$ on $Z$ using data set $I_{j1}$, and then applies the estimated model and uses $Z$ in data set $I_{j2}$ to predict $Y_{j1}$ where $j_1, j_2 = 1, 2$ and $j_1 \neq j_2$. The predicted values of $Y_{j1}$ are then treated as true values, which together with the observations of $(Y_{j2,i}', Z_i')$ in the data set $I_{j2}$ are used in estimating $\theta_0$. Such a data imputed estimator, however, is not guaranteed to work well in practice. For general nonlinear models, they are usually inconsistent.

In this paper, we consider a different approach. It is motivated by the moment-matching estimators proposed in [Klevmarken (1982), Angrist and Krueger (1992), Arellano and Meghir (1992) and Ridder and Moffitt (2007)]. However, instead of matching unconditional moments, we match the conditional moments in (1) for estimating $\theta_0$ and consider optimality in the class of conditional moment-matching estimators. Let $\phi_j(Z, \theta) = E[g_j(Y_{j, \theta})|Z]$ for $j = 1, 2$ and $\phi(Z, \theta) = \phi_1(Z, \theta) - \phi_2(Z, \theta)$. We can write the conditional moment condition in (1) as

$$\phi(Z, \theta) = \phi_1(Z, \theta_0) - \phi_2(z, \theta_0) = 0_{ds}. \quad (2)$$

Since $(Y_{j1}', Z')$ are jointly observed, the conditional mean function $\phi_j(z, \theta)$ is identified and can be consistently estimated using the data set $I_j (j = 1, 2)$. Therefore a consistent estimator of $\theta_0$ can be constructed using the semiparametric minimum distance (SMD) estimation which minimizes the distance between the estimators of functions $\phi_j(\cdot, \theta) (j = 1, 2)$ with respect to $\theta$.

As the IVs $Z$ are available in both data sets, we have $n \times (n = n_1 + n_2)$ observations on $Z$: $\{Z_i\}_{i \in I}$ where $I = I_1 \cup I_2$. Let $P_k(z) = (p_1(z), \ldots, p_k(z))'$ be a $k$-dimensional vector of approximating functions (such as power series, splines, Fourier series, etc.) for any $k \in \mathbb{N}$. For any $k, n \in \mathbb{N}$, we define $P_{n,k} = ((P_k(Z_i))_{i \in I})'$ which is an $n \times k$ matrix. Accordingly, we define $P_{n,j,k} = ((P_k(Z_i))_{i \in I_j})$ for $j = 1, 2$, which is an $n_j \times k_j$ matrix. The conditional mean function $\phi_j(z, \theta)$ is estimated by the series estimation

$$\hat{\phi}_j(z, \theta) = P_{k_j}(z)'(P_{n_j,k_j}P_{n_j,k_j})^{-1}P_{n_j,k_j}g_{j,n_j}(\theta) \quad (3)$$

where $g_{j,n_j}(\theta) = ((g_j(Y_{j,i}, \theta))_{i \in I_j})'$. The set of approximating functions $P_k(z)$ used in estimating $\phi_j(z, \theta)$ do not need to be the same, but we suppress its dependence on $j$ for the simplicity of nota-

---

See the simulation results on the $Y_1$-imputed estimator and the $Y_2$-imputed estimator in Section 8 for simple illustrations.
tions. Using the nonparametric estimator of $\phi_j(z, \theta)$, we can estimate $\theta_0$ via the SMD estimation:

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left[ \hat{\phi}(Z_i, \theta)' \hat{W}_n(Z_i) \hat{\phi}(Z_i, \theta) \right]$$  \hspace{1cm} (4)

where $\hat{\phi}(Z_i, \theta) = \hat{\phi}_1(Z_i, \theta) - \hat{\phi}_2(Z_i, \theta)$, $\Theta$ denotes the parameter space containing $\theta_0$, $\hat{W}_n(\cdot)$ is a $d_g \times d_g$ symmetric matrix function of $z \in \mathcal{Z}$. The simplest and the most straightforward choice of $\hat{W}_n(\cdot)$ is the identity matrix, i.e., $\hat{W}_n(z) = I_{d_g}$ for any $z \in \mathcal{Z}$. We denote this identity matrix weighted SMD estimator $\tilde{\theta}_n$ and name it the preliminary estimator of $\theta_0$. As we show below, the preliminary estimator may not have the smallest possible asymptotic variance. However, it is useful to construct the optimal weight function for a more efficient SMD estimator.

The SMD estimator $\hat{\theta}_n$ looks similar to the sieve minimum distance estimator studied in Ai and Chen (2003), Chen and Pouzo (2009), Chen and Pouzo (2012) and Chen and Pouzo (2015). The key difference is that in the SMD estimation adopted here, the conditional mean function $\phi(z, \theta)$ is estimated through the estimations of $\phi_1(z, \theta)$ and $\phi_2(z, \theta)$ separately, using the corresponding data sets $I_j$ for $j = 1, 2$. In contrast, the in the sieve minimum distance estimation of Ai and Chen (2003), for example, $(Y_1, Y_2, Z)$ are jointly observed and hence $\phi(z, \theta)$ is estimated using only one sample. This separate estimation of $\phi_j(z, \theta)$ ($j = 1, 2$) requires one to take into account the estimation errors in $\phi_j(z, \theta)$ ($j = 1, 2$) separately. In turn, both the finite sample and the asymptotic variance-covariance matrices of the SMD estimator $\hat{\theta}_n$ are different from the case in which $\phi(z, \theta)$ can be estimated using only one sample.

Another way of utilizing the conditional moment conditions (1) and the two data sets $I_1$ and $I_2$ is to consider a set of unconditional moment conditions

$$E[(g_1(Y_1, \theta_0) - g_2(Y_2, \theta_0)) \otimes P_k(Z)] = E[g_1(Y_1, \theta_0) \otimes P_k(Z)] - E[g_2(Y_2, \theta_0) \otimes P_k(Z)] = 0_{k \times d_g},$$

which are implied by the conditional moment conditions in (1), and then estimate $\theta_0$ by the two-sample GMM estimation proposed in Ridder and Moffitt (2007). The statistical properties of the two-sample GMM estimator are only justified in Ridder and Moffitt (2007) in the case where $k$ is a fixed number. However, in order to fully use the information of the conditional moment conditions (1) and ensure consistent estimation of $\theta_0$, one must to let $k$ goes to infinity as the the sample size increases. While it would of interest to study the statistical properties of the two-sample GMM estimator with many unconditional moment conditions, it is beyond the scope of our current study.

3 Asymptotic Properties of the SMD Estimator

In this section, we establish the asymptotic properties of the SMD estimator $\hat{\theta}_n$. First we state the sufficient conditions for the consistency of $\hat{\theta}_n$.  

6
Assumption 1. (i) \{ (Y'_{1i}, Z'_{1i}) \}_{i \in I_1} and \{ (Y'_{2i}, Z'_{2i}) \}_{i \in I_2} are random samples which are independent of each other; (ii) for \( j = 1, 2 \), \( \sup_{\theta \in \Theta} E \left[ \| g_j(Y_j, \theta) \|^2 \right] < C \) and
\[
\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_j(Z_i, \theta) - \phi_j(Z_i, \theta) \right\|^2 = o_p(1);
\]
(iii) for any \( \varepsilon > 0 \), there exists \( \eta_{\varepsilon} > 0 \) such that
\[
\inf_{\{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon\}} E \left[ \| \phi(Z, \theta) \|^2 \right] > \eta_{\varepsilon};
\]
(iv) \( \sup_{z \in \mathcal{Z}} \| \hat{W}_n(z) - W_n(z) \| = O_p(\delta_{w,n}) \) where \( \delta_{w,n} = o(1) \), and \( W_n(\cdot) \) is a sequence of symmetric non-random matrix function which satisfies
\[
C^{-1} \leq \inf_{z \in \mathcal{Z}} \lambda_{\min}(W_n(z)) \leq \sup_{z \in \mathcal{Z}} \lambda_{\max}(W_n(z)) \leq C, \text{ for any } n;
\]
(v) \( \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \| \phi(Z_i, \theta) \|^2 - E[\| \phi(Z, \theta) \|^2] = o_p(1) \).

Assumption 1(i) requires that the data are i.i.d. and the two data sets are independent of each other. Assumption 1(ii) imposes a uniform finite second moment bound on \( g_j(Y_j, \theta) \) \( (j = 1, 2) \). It also requires that the nonparametric estimators \( \hat{\phi}_j(Z_i, \theta) \) of \( \phi_j(Z_i, \theta) \) \( (j = 1, 2) \) are consistent under the empirical \( L_2 \)-norm uniformly over \( \theta \in \Theta \). The uniform consistency of \( \hat{\phi}_j(Z_i, \theta) \) can be proved under low level sufficient conditions (see, e.g., Corollary A.1 in [Ai and Chen (2003)] for verification of \( \text{ in a more complicated case where } \theta \text{ may includes infinite dimensional parameters} \). Assumption 1(iii) is the identification condition for \( \theta_0 \). Assumption 1(iv) implies that \( \hat{W}_n(\cdot) \) is a consistent estimator of the nonrandom matrix function \( W_n(\cdot) \) uniformly over \( z \). It is clear that Assumption 1(iv) holds trivially if \( \hat{W}_n(\cdot) \) is the identity matrix. Assumption 1(v) is a uniform law of large numbers on the random functions \( \| \phi(Z, \theta) \|^2 \) indexed by \( \theta \in \Theta \), which can also be verified under low level sufficient conditions (see, e.g., Corollary A.2(i) in [Ai and Chen (2003)]).

Theorem 1. Under Assumption 1, we have \( \hat{\theta}_n = \theta_0 + o_p(1) \).

For \( j = 1, 2 \) and \( m = 1, \ldots, d_y \), let \( g_{j,m}(y_j, \theta) \) and \( \phi_{j,m}(z, \theta) \) denote the \( m \)-th component in \( g_j(y_j, \theta) \) and \( \phi_j(z, \theta) \) respectively. For ease of notation, we define
\[
\begin{align*}
g_j^{(1)}(Y_j, \theta) &= \partial g_j(Y_j, \theta) / \partial \theta', \\
g_j^{(2)}(Y_j, \theta) &= \partial^2 g_j(Y_j, \theta) / \partial \theta \partial \theta', \\
\phi^{(1)}(z, \theta) &= \partial \phi(z, \theta) / \partial \theta', \\
u_j &= g_j(Y_j, \theta_0) - \phi_j(Z, \theta_0), \\
Q_k &= E [ P_k(Z) P_k(Z)' ], \\
\Sigma_j(Z) &= E [ u_j u_j' | Z ], \\
H_n &= E \left[ \phi^{(1)}(Z, \theta_0)' W_n(Z) \phi^{(1)}(Z, \theta_0) \right], \\
Q_{n,j,u_j} &= n_j^{-1} \sum_{i \in I_j} \Sigma_j(Z_i) \otimes (P_{k_j}(Z_i) P_{k_j}'(Z_i)).
\end{align*}
\]

We next state the sufficient conditions for the asymptotic normality of \( \hat{\theta}_n \).
Assumption 2. The following conditions hold:

(i) \( \max_{j=1,2} \sup_{\theta \in \Theta} n_j^{-1} \sum_{i \in I_j} \sum_{m=1}^{d_j} ||g_{j,m}^{(2)}(Y_j, \theta)||^2 = O_p(1) \);
(ii) \( \lambda_{\min}(H_n) \geq C^{-1} \) for any \( n \);
(iii) \( E \left[ ||g_j^{(1)}(Y_j, \theta_0)||^4 \right] + \sup_{z \in Z} E \left[ ||g_j^{(1)}(Y_j, \theta_0)||^2 \right] Z = z < C \) for \( j = 1, 2 \);
(iv) \( C^{-1} \leq \lambda_{\min}(Q_k) \leq \lambda_{\max}(Q_k) \leq C \) for any \( k \in \mathbb{N} \);
(v) \( C^{-1} \leq \min_{j=1,2} \inf_{z \in Z} \lambda_{\min}(\Sigma_j(z)) \) and \( \max_{j=1,2} \sup_{z \in Z} E[||u_j||^4 | Z = z] \leq C \).

Assumption 2(i) holds if \( \sum_{m=1}^{d_j} ||g_{j,m}^{(2)}(y_j, \theta)||^2 \leq C \) for any \( y_j \) in the support of \( Y_j \) and for any \( \theta \) in the parameter space \( \Theta \). It can also be verified by the uniform moment bound

\[
\max_{j=1,2} \sup_{\theta \in \Theta} E \left[ \sum_{m=1}^{d_j} ||g_{j,m}^{(2)}(Y_j, \theta)||^2 \right] \leq C
\]

and the uniform law of large numbers on \( \sum_{m=1}^{d_j} ||g_{j,m}^{(2)}(Y_j, \theta)||^2 \) over \( \theta \in \Theta \). The lower bound of the eigenvalues of \( H_n \) in Assumption 2(ii) ensures the local identification of \( \theta_0 \). Assumption 2(iii) imposes a finite fourth moment bound and a conditional second moment bound on \( g_j^{(1)}(Y_j, \theta_0) \). Assumption 2(iv) is a regularity condition in the series estimation literature, see e.g., Andrews [1991], Newey [1997] and Chen [2007]. Assumption 2(v) imposes a finite conditional fourth moment bound on \( u_j \) and a uniform non-singularity condition on the conditional variance-covariance \( \Sigma_j(z) \).

Assumption 3. (i) For any function \( a(z) \) with \( E[|a(Z)|^2] \leq C \), there are \( k \times 1 \) vectors \( \beta_k \) such that as \( k \to \infty \),

\[
E \left[ a(Z) - P_k(Z)' \beta_k \right] \to 0;
\]

(ii) for \( j = 1, 2 \) and \( m = 1, \ldots, d_j \), there exist \( \beta_{\phi_j,m,k} \in \mathbb{R}^k \) and \( r_\phi > 1 \) such that

\[
E \left[ ||\phi_{j,m}(Z, \theta_0) - P_k(Z)' \beta_{\phi_j,m,k}||^2 \right] = O(k^{-r_\phi});
\]

(iii) \( \max_{j=1,2}(\xi_k^2 \log(k_j) + k_j^{1/2}n_j^{-1/2} + n_j^{1/2}k_j^{-r_\phi} + \delta_w n_j^{1/2}) = o(1) \) where \( \xi_k = \sup_{z \in Z} \|P_k(z)\| \).

Assumption 3(i) requires that the functions \( P_k(z) \) approximate any square integrable function \( a(z) \) well. By Assumption 3(ii) and Jensen’s inequality

\[
E \left[ ||\phi_{j,m}(Z, \theta_0)||^2 \right] \leq E \left[ ||g_{j,m}(Z, \theta_0)||^2 \right] < C.
\]

This, together with Assumptions 3(i) and 3(iii) implies that there exists \( \beta_{\phi_j,m,k} \in \mathbb{R}^k \) such that

\[
E \left[ ||\phi_{j,m}(Z, \theta_0) - P_k(Z)' \beta_{\phi_j,m,k}||^2 \right] = o(1).
\]

Assumption 3(ii) further implies that the above approximation error is of the order \( k^{-r_\phi} \). The constant \( r_\phi \) is related to the smoothness of \( \phi_m(z, \theta_0) \) in terms of \( z \) for any \( m = 1, \ldots, d_g \). Assumption
For the preliminary estimator $\tilde{\theta}_n$ based on the identity weight matrix, $\delta_{w,n} = 0$ and the above condition (6) holds if $k_j = O(n_j^2)$ for any $\alpha \in (1/(2r_\phi), 1/2)$. For the SMD estimators based on non-identity weight matrices, further restrictions on $k_j$ may be imposed through the condition that $\max_{j=1,2} \delta_{w,n}k_j^{1/2} = o(1)$.

Let $W_{n,m}(\cdot)$ denote the $m$-th column of $W_n(\cdot)$ and define $\phi_{w\theta,m,n} \equiv (\phi(1)(Z_i, \theta_0)W_{n,m}(Z_i))_{i \in I}$ for $m = 1, \ldots, d_g$. It is clear that $\phi_{w\theta,m,n}$ is a $d_\theta \times n$ matrix for $m = 1, \ldots, d_g$. Let $\phi_{w\theta,n} \equiv (\phi_{w\theta,1,n}, \ldots, \phi_{w\theta,d_g,n})$. It is clear that $\phi_{w\theta,n}$ is a $d_\theta \times (nd_g)$ matrix. For $j = 1, 2$, define

$$\Omega_{j,n} \equiv \phi_{w\theta,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})Q_{n_j,u_j}(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j})\phi_{w\theta,n},$$

where $Q_{n_j,k_j} = n_j^{-1} \sum_{i \in I_j} P_{k_j}(Z_i)P'_{k_j}(Z_i)$. The matrix $\Omega_{j,n}$ is the variance-covariance matrix of the (first-order) estimation error in $\tilde{\theta}_n$ introduced by estimating $\phi_j(z, \theta_0)$ with sample $I_j$.

**Theorem 2.** Under Assumptions 2 and 3, we have

$$\tilde{\theta}_n - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2})$$

and moreover

$$\gamma'_n(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{1/2}(\tilde{\theta}_n - \theta_0) \rightarrow_d N(0,1)$$

for any non-random sequence $\gamma_n \in R^{d_\theta}$ with $\gamma'_n\gamma_n = 1$.

**Remark 1.** The first result of Theorem 2, i.e., (7), implies that the convergence rate of the SMD estimator is of the order $\max\{n_1^{-1/2}, n_2^{-1/2}\}$. The asymptotic normality of the SMD estimator is derived without requiring the ratio of the two sample sizes (say $n_1/n_2$) to be convergent. In the extreme case that $n_1/n_2 \rightarrow 0$ (or $n_2/n_1 \rightarrow 0$), the asymptotic normality of the SMD estimator takes a simplified form. By (13) in the Appendix, the smallest eigenvalues of $n_j\Omega_{j,n}$ ($j = 1, 2$) are bounded away from zero with probability approaching 1. Therefore, if $n_1/n_2 \rightarrow 0$,

$$n_1^{-1}H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n = H_n\left(n_1\Omega_{1,n} + \frac{n_1}{n_2}n_2\Omega_{2,n}\right)^{-1}H_n = H_n(n_1\Omega_{1,n})^{-1}H_n + o_p(1)$$
which together with (7), Assumption 2(ii) and the Slutsky Theorem implies that

\[
\gamma'_n(H_n\Omega^{-1}_{1,n}H_n)^{1/2}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, 1).
\]

Similar result can be established in the case that \(n_2/n_1 \rightarrow 0\).

**Remark 2.** By Theorem 2 and the Cramér-Wold Theorem

\[
(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{1/2}(\hat{\theta}_n - \theta_0) \rightarrow_d N(\mathbf{0}_{d_\theta}, I_{d_\theta}),
\]

which together with the continuous mapping theorem (CMT) implies that,

\[
(\hat{\theta}_n - \theta_0)'(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)(\hat{\theta}_n - \theta_0) \rightarrow_d \chi^2(d_\theta).
\]

Moreover, let \(\iota_j\) be the \(d_\theta \times 1\) selection vector whose \(j\)-th \((j = 1, \ldots, d_\theta)\) component is 1 and rest components are 0. Define

\[
\gamma_{j,n} = \frac{(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{-1/2}}{(\iota_j'H_n^{-1}(\Omega_{1,n} + \Omega_{2,n})H^{-1}_n\iota_j)^{1/2}} \iota_j, \text{ for } j = 1, \ldots, d_\theta.
\]

It is clear that \(\gamma'_j, \gamma_{j,n} = 1\). By Theorem 2, we have

\[
\gamma_{j,n}(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{1/2}(\hat{\theta}_n - \theta_0) = \frac{\hat{\theta}_{j,n} - \theta_{j,0}}{(\iota_j'H_n^{-1}(\Omega_{1,n} + \Omega_{2,n})H^{-1}_n\iota_j)^{1/2}} \rightarrow_d N(0, 1),
\]

where \(\hat{\theta}_{j,n} = \iota_j'\hat{\theta}_n\) and \(\theta_{j,0} = \iota_j'\theta_0\). Results in \([10]\) and \([11]\) can be used to conduct inference on \(\theta_0\) and \(\theta_{j,0}\), if consistent estimators of \(H_n, \Omega_{1,n}\) and \(\Omega_{2,n}\) are available.

**Remark 3.** Estimators of \(H_n, \Omega_{1,n}\) and \(\Omega_{2,n}\) can be constructed using their empirical counterparts. First, the matrix \(H_n\) can be estimated by

\[
\hat{H}_n \equiv n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \hat{\theta}_n)'\hat{W}_n(Z_i)\hat{\phi}^{(1)}(Z_i, \hat{\theta}_n)
\]

where \(\hat{\phi}^{(1)}(Z_i, \hat{\theta}_n) \equiv \hat{\phi}_1^{(1)}(Z_i, \hat{\theta}_n) - \hat{\phi}_2^{(1)}(Z_i, \hat{\theta}_n)\) for any \(i \in I\). To construct the estimator of \(\Omega_{j,n}\), we define

\[
\hat{Q}_{n_j,u_j} \equiv n_j^{-1} \sum_{i \in I_j} \hat{u}_{j,i} \hat{u}_{j,i}' \otimes (P_{k_i}(Z_i)P_{k_j}(Z_i))
\]
where \( \hat{u}_{j,i} \equiv g_j(Y_{j,i}, \hat{\theta}_n) - \hat{\phi}_j(Z_i, \hat{\theta}_n) \) for any \( i \in I_j \), and

\[
\hat{\phi}_{w\theta,n} \equiv (\hat{\phi}_{w\theta,1,n}, \ldots, \hat{\phi}_{w\theta,d_g,n})
\] (14)

where \( \hat{\phi}_{w\theta,m,n} \equiv (\hat{\phi}^{(1)}(Z_i, \hat{\theta}_n)W_n(m(z_i)))_{i \in I} \) for any \( m = 1, \ldots, d_g \). The matrix \( \Omega_{j,n} \) can be estimated by

\[
\hat{\Omega}_{j,n} \equiv \frac{\hat{\phi}_{w\theta,n}(I_{d_g} \otimes P_{n,k_j}Q^{-1}_{n,j,k_j})\hat{Q}_{n,j,n}(I_{d_g} \otimes Q^{-1}_{n,j,k_j}P'_{n,k_j})\hat{\phi}'_{w\theta,n}}{n^2n_j},
\] (15)

for \( j = 1, 2 \). The estimator of the variance-covariance matrix of the SMD estimator can be constructed as

\[
\hat{V}_n = \hat{H}_n^{-1}(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})\hat{H}_n^{-1}.
\] (16)

The estimators \( \hat{H}_n \) and \( \hat{\Omega}_{j,n} \) involve the weight function \( \hat{W}_n(z) \), which is the identity matrix for the preliminary SMD estimator. We discuss the optimal choice of the weight function and its consistent estimator in the next section.

4 Optimal Weighting

In this section, we compare the SMD estimators through their finite sample variances. The comparison leads to an optimal matrix of weight functions which produces SMD estimator with the smallest finite sample variance, as well as asymptotic variance, among all SMD estimators. The following lemma simplifies the finite sample variance-covariance matrix which facilitates the comparison of the SMD estimators.

**Lemma 1.** Under Assumptions [7](i), [7](ii), [7](iv), [2](ii), [2](iv), [3](v) and [3](iii),

\[
H_n^{-1}(\Omega_{1,n} + \Omega_{2,n})H_n^{-1} = V_n(1 + o_p(1)).
\]

where

\[
V_n \equiv H_n^{-1}\left( E \left[ \phi^{(1)}(Z, \theta_0)W_n(Z) \left( \frac{\Sigma_1(Z)}{n_1} + \frac{\Sigma_2(Z)}{n_2} \right) W_n(Z)\phi^{(1)}(Z, \theta_0) \right] \right) H_n^{-1}.
\]

We call \( H_n^{-1}(\Omega_{1,n} + \Omega_{2,n})H_n^{-1} \) and \( V_n \) the finite sample variance and the pre-asymptotic variance of the SMD estimator \( \hat{\theta}_n \) with weight function \( W_n(\cdot) \), respectively. Since the finite sample variance-covariance matrix of the SMD estimator can be approximated by \( V_n \), comparison of the SMD estimators is conducted through \( V_n \). If we set the weight matrix to

\[
W^*_n(z) \equiv (n_1^{-1} + n_2^{-1})(n_1^{-1}\Sigma_1(z) + n_2^{-1}\Sigma_2(z))^{-1},
\] (17)
then the pre-asymptotic variance of the SMD estimator becomes

\[ V_n^* \equiv (n_1^{-1} + n_2^{-1})(H_n^*)^{-1} \]  

(18)

where \( H_n^* \equiv (n_1^{-1} + n_2^{-1})E \left[ \phi^{(1)}(Z, \theta_0)'(n_1^{-1}\Sigma_1(Z) + n_2^{-1}\Sigma_2(Z))^{-1}\phi^{(1)}(Z, \theta_0) \right] \). The next lemma shows that \( V_n^* \) is the smallest variance-covariance among all SMD estimators.

**Theorem 3.** For weight function \( W_n(\cdot) \) satisfying Assumption \( \square(iv) \), we have \( V_n \geq V_n^* \) for any \( n_1 \) and any \( n_2 \).

We call the SMD estimators whose pre-asymptotic variance-covariance matrices equal \( V_n^* \) optimal SMD estimators. To construct the optimal SMD estimator, we have to get the estimator of the optimal weight \( W_n^*(z) \) and show that Assumption \( \square(iv) \) holds. Since \( \Sigma_1(z) \) and \( \Sigma_2(z) \) are symmetric matrix functions of \( z \), \( W_n^*(z) \) is also a symmetric matrix function. The next lemma shows that \( W_n^*(z) \) satisfies the eigenvalue restrictions imposed in Assumption \( \square(iv) \).

**Lemma 2.** Under Assumption \( \square(v) \),

\[ C^{-1} \leq \inf_{z \in \mathcal{Z}} \lambda_{\min}(W_n^*(z)) \leq \sup_{z \in \mathcal{Z}} \lambda_{\max}(W_n^*(z)) \leq C \]

for any \( n_1 \) and any \( n_2 \).

For feasible optimal SMD estimation, one needs to estimate the optimal weight function \( W_n^*(\cdot) \) and derive the convergence rate of the estimator. Since the unknown components in \( W_n^*(z) \) are the conditional variance-covariance matrices \( \Sigma_j(z) \) for \( j = 1, 2 \), an estimator of \( W_n^*(z) \) can be constructed using the estimators of \( \Sigma_j(z) \), \( j = 1, 2 \). Recall that \( \tilde{\theta}_n \) denotes the preliminary SMD estimator with the identity weight function. For \( j = 1, 2 \), define \( \tilde{u}_{j,i} = g_j(Y_{j,i}, \tilde{\theta}_n) - \tilde{\phi}_j(Z_i, \tilde{\theta}_n) \) for any \( i \in I_j \), and for \( m = 1, \ldots, d_g \) let \( \tilde{u}_{j,m,i} \) denote the \( m \)-th component in \( \tilde{u}_{j,i} \). The estimator of \( \Sigma_j(z) \) is denoted by \( \tilde{\Sigma}_j(z) \), whose \((m_1, m_2)\)-th component is

\[ \tilde{\Sigma}_{j,m_1,m_2}(z) \equiv P_{k_j}^t(z)(P_{n_j,k_j}^t P_{n_j,k_j})^{-1}P_{n_j,k_j}^t \tilde{U}_{j,m_1,m_2,n_j}, \]

(19)

where \( \tilde{U}_{j,m_1,m_2,n_j} = ((\tilde{u}_{j,m_1,i}\tilde{u}_{j,m_2,i})_{i \in I_j})' \). The optimal weight matrix is then estimated by

\[ \hat{W}_n^*(z) \equiv (n_1^{-1} + n_2^{-1})(n_1^{-1}\tilde{\Sigma}_1(z) + n_2^{-1}\tilde{\Sigma}_2(z))^{-1}. \]

(20)

Feasible optimal SMD estimation is carried out by replacing \( \hat{W}_n(z) \) in \( (4) \) with \( \hat{W}_n^*(z) \).

5 **Consistency of the Variance-Covariance Estimator and the Empirical Optimal Weight**

In this section, we prove the consistency of the variance-covariance matrix estimator \( \hat{V}_n \) for the SMD estimator, and also the empirical optimal weight function \( \hat{W}_n^*(z) \) constructed in the previous sections. To do this, the following conditions are needed.
Assumption 4.  
(i) $\max_{j=1,2} \sup_{\theta \in \Theta} n_j^{-1} \sum_{i \in I_j} ||g_j^{(1)}(Y_{j,i}, \theta)||^4 = O_p(1)$;  
(ii) for any $m_1, m_2 = 1, \ldots, d_g$, there exist $\beta_{\Sigma_j, m_1, m_2, k} \in \mathbb{R}^k$ and $r_u > 0$ such that

$$\sup_{z \in \mathcal{Z}} |\Sigma_{j, m_1, m_2}(z) - P_k(z)' \beta_{\Sigma_j, m_1, m_2, k}| = O(k^{-r_u});$$

(iii) $\max_{j=1,2} \left( \xi_k k_j^{1/2}(n_1^{-1/2} + n_2^{-1/2}) + \xi_k k_j^{-r_u} \right) = o(1).$

Assumption 4(i) is similar to Assumption 2(i). It holds when $g_j^{(1)}(y_j, \theta)$ is a bounded function for any $y_j$ in the support of $Y_j$ and for any $\theta \in \Theta$. In general, one can verify Assumption 4(ii) using $\sup_{\theta \in \Theta} E[||g_j^{(1)}(Y_j, \theta)||^4] < C$ and the following uniform law of large numbers for $j = 1, 2$,

$$\sup_{\theta \in \Theta} \left\{ n_j^{-1} \sum_{i \in I_j} ||g_j^{(1)}(Y_{j,i}, \theta)||^4 - E[||g_j^{(1)}(Y_j, \theta)||^4] \right\} = o_p(1).$$

Assumption 4(iii) implies that the conditional variance-covariance matrices $\Sigma_j(z)$ can be well approximated by $P_k(z)$ with approximation error of the order $k^{-r_u}$. Assumption 4(iii) imposes further restrictions on the numbers of the approximating functions.

**Theorem 4.** Suppose Assumptions 1, 2, 3 and 4 hold. We have

$$\tilde{H}_n(\tilde{\Omega}_{1,n} + \tilde{\Omega}_{2,n})^{-1} \tilde{H}_n = (I_{d_\theta} + o_p(1)) H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} H_n$$

(21)

and moreover,

$$\gamma_n'(\tilde{H}_n(\tilde{\Omega}_{1,n} + \tilde{\Omega}_{2,n})^{-1} \tilde{H}_n)^{1/2}(\tilde{\theta}_n - \theta_0) \rightarrow_d N(0, 1),$$

(22)

for any non-random sequence $\gamma_n \in \mathbb{R}^{d_{\theta}}$ with $\gamma_n'\gamma_n = 1$.

**Remark 4.** By the Cramér-Wold Theorem and (22) in Theorem 4,

$$(\tilde{H}_n(\tilde{\Omega}_{1,n} + \tilde{\Omega}_{2,n})^{-1} \tilde{H}_n)^{1/2}(\tilde{\theta}_n - \theta_0) \rightarrow_d N(0, I_{d_\theta}),$$

which together with the CMT implies that

$$(\tilde{\theta}_n - \theta_0)'(\tilde{H}_n(\tilde{\Omega}_{1,n} + \tilde{\Omega}_{2,n})^{-1} \tilde{H}_n)(\tilde{\theta}_n - \theta_0) \rightarrow_d \chi^2(d_{\theta}).$$

(23)

By the consistency of the variance-covariance matrix estimator, i.e., (21) in Theorem 4, we have

$$\epsilon_j' \tilde{H}_n(\tilde{\Omega}_{1,n} + \tilde{\Omega}_{2,n})^{-1} \tilde{H}_n \epsilon_j = (1 + o_p(1)) \epsilon_j' H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} H_n \epsilon_j$$

(24)

where $\epsilon_j$ is the $d_{\theta} \times 1$ selection vector whose $j$-th ($j = 1, \ldots, d_{\theta}$) component is 1 and rest components
are 0. By (11), (24) and the CMT,
\[ t_{j,n}(\theta_{j,0}) = \frac{\hat{\theta}_{j,n} - \theta_{j,0}}{\left( \left( I_j \hat{H}_n^{-1} \Omega_{1,n} + \hat{H}_n^{-1} t_j \right) \right)^{1/2}} \rightarrow_d N(0, 1). \] (25)

The Wald-statistic in (23) and the Student-t statistic in (25) can be applied to conduct joint inference on \( \theta_0 \) and inference on \( \theta_{j,0} \) for \( j = 1, \ldots, d \), respectively. Moreover, one can calculate the standard error estimator of the SMD estimator using (24) directly.

Using the empirical optimal weight function \( \hat{W}_{n}^*(z) \) given in (20), one can construct the following optimal SMD estimator
\[ \hat{\theta}_n^* = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i \in I} \left[ \Delta \hat{\phi}(Z_i, \theta) \hat{W}_{n}^*(Z_i) \Delta \hat{\phi}(Z_i, \theta) \right]. \] (26)

To show the optimality of \( \hat{\theta}_n^* \), it is sufficient to show that \( \hat{W}_{n}^*(Z_i) \) satisfies the high level conditions in Assumption 1(iv). Since we have shown in Lemma 2 that the population version of \( \hat{W}_{n}^*(z) \), i.e. \( W_n^*(z) \), satisfies the eigenvalue restrictions in Assumption 1(iv), we only need to derive the convergence rate of \( \hat{W}_{n}^*(z) \) and make sure that the rate requirement in Assumption 2(iv) is satisfied.

**Theorem 5.** Under Assumptions 1, 2, 3 and 4, we have
\[ \sup_{z \in \mathcal{Z}} \left| \hat{W}_n^*(z) - W^*(z) \right| = O_p(\delta_{w,n}), \]
where \( \delta_{w,n} = \max_{j=1,2} (\xi_{k_j} (k_1^{1/2} + k_2^{1/2}) n_j^{-1/2} + n_1^{1/2} + n_2^{1/2}) \).

**Remark 5.** By Assumption 4(iii) and Theorem 5, the empirical optimal weight function \( \hat{W}_{n}^*(z) \) is a uniform consistent estimator of the optimal weight function \( W_n^*(z) \). The convergence rate \( \delta_{w,n} \) has to satisfy \( (k_1^{1/2} + k_2^{1/2}) \delta_{w,n} \) in Assumption 1(iv). To simplify the discussion, we take \( k_1 = k_2 = k \). Then the rate requirement is satisfied if
\[ \xi_k k^{1/2} (k^{1/2} + k^{1/2}) n_j^{-1/2} + n_1^{1/2} + n_2^{1/2} = o(1). \] (27)

When the power series are used as the approximating functions \( P_k(z) \), we have \( \xi_{k_j} \leq Ck_j \). In this case, (27) becomes to
\[ k^{3/2-r_u} + k^{2} n_1^{-1/2} + n_2^{1/2} = o(1) \]
which holds when \( r_u > 3/2 \) and \( k = o(n_1^{-1/4} + n_2^{-1/4}) \). When the splines or trigonometric functions are used as the approximating functions \( P_k(z) \), we have \( \xi_{k_j} \leq Ck_j^{1/2} \). Then the rate restriction in
Assumption (iv) becomes
\[ k^{1-r_u} + k^{3/2}(n_1^{-1/2} + n_2^{-1/2}) = o(1), \]
which holds when \( r_u > 1 \) and \( k = o(n_1^{-1/3} + n_2^{-1/3}) \) and is slightly weaker than the corresponding conditions for the power series.

6 Monte Carlo Simulation

In this section, we study the finite sample performances of the SMD estimator and the proposed inference method. The simulated data is from the following model

\[ Y_{1,i} = g_2(Y_{2,i}, \theta_0) + v_i, \quad (28) \]

where \( \theta_0 = 1 \) is the unknown parameter, \( Y_{1,i}, Y_{2,i} \) and \( v_i \) are random variables, \( g_2(\cdot, \cdot) \) is a function specified in the following

\[ g_2(y_2, \theta) = \begin{cases} y_2 \theta & \text{in Model 1} \\ \log(1 + y_2^2 \theta) & \text{in Model 2} \end{cases}. \quad (29) \]

Let \((X^*_{1,i}, X^*_{2,i}, v^*_i)\) be a standard normal random vector for any \( i \). The instrumental variable \( Z_i \), the regressor \( Y_{2,i} \) and the regression error \( v_i \) are generated in the following way

\[ Z_i = (1 + X^*_{2,i})(4 + 4X^*_{2,i}^2)^{-1/2}, \quad Y_{2,i} = Z_i + Z_i^2 X^*_{1,i} \quad \text{and} \quad v_i = Y_{2,i}^2 v^*_i. \quad (30) \]

We assume that \((Y_{1,i}, Z_i)\) are observed together, and \((Y_{2,i}, Z_i)\) are observed together. The simulation model is a much simplified version of the general model studied in the paper since here we have \( g_1(Y_{1,i}, \theta_0) = Y_{1,i} \), and \( g_1(\cdot, \cdot) \) and \( g_2(\cdot, \cdot) \) are scalar functions.

We generate i.i.d. random vectors \((X^*_{1,i}, X^*_{2,i}, v^*_i)\) \((i = 1, \ldots, n_1 + n_2)\) from the standard normal distribution and then calculate \((Y_{1,i}, Y_{2,i}, Z_i)\) \((i = 1, \ldots, n_1 + n_2)\) using equations (28) and (30). The first data set \(\{(Y_{1,i}, Z_i)\}_{i \in I_1}\) and the second data set \(\{(Y_{2,i}, Z_i)\}_{i \in I_2}\) are then drawn from \(\{(Y_{1,i}, Y_{2,i}, Z_i)\}_{i=1}^n\) by setting \(I_1 = \{1, \ldots, n_1\}\) and \(I_2 = \{n_1 + 1, \ldots, n_1 + n_2\}\). By the data generating process, the two data sets are independent of each other. As both the magnitudes of \(n_1, n_2\) and their relative magnitudes are important to the finite sample properties of the SMD estimator, we consider two sampling schemes: equal sampling and unequal sampling, separately. In the equal sampling scheme, we set \(n_1 = n_2 = n_0\) where \(n_0\) starts from 50 with increments of 50 up to and ends at 1,000. In the unequal sampling, we set \(n_1 + n_2 = 1,000\) where \(n_1\) starts from 100 with increment 50 and ends at 900. For each combination of \(n_1\) and \(n_2\), we generate 10,000 simulated samples to evaluate the performances of the SMD estimator and the proposed inference procedure.

In addition to the SMD estimator, we study two alternative estimators based on data impu-
The first estimator (which is called the $Y_2$-imputed estimator in this section) is defined as
\[
\hat{\theta}_{Y_2,n} = \arg \min_{\theta \in \Theta} \frac{1}{n_1} \sum_{i \in I_1} (Y_{1,i} - g(\hat{Y}_{2,i}, \theta))^2
\] (31)
Figure 2: Properties of the MD and the Imputation Estimators \((n_1 + n_2 = 1,000)\)

\[
\hat{Y}_{2,i} = n_2^{-1} P_{k_2}^T (Z_i) Q_{n_2,k_2}^{-1} \sum_{i \in I_2} Y_{2,i} P_{k_2} (Z_i)
\]

for any \(i \in I_1\) is the predicted value of \(Y_{2,i}\) in the first data set based on a nonparametric regression using the second data set. The second estimator
(which is called the $Y_1$-imputed estimator in this section) is defined as

$$
\hat{\theta}_{Y_1,n} = \arg \min_{\theta \in \Theta} \sum_{i \in I_2} (\hat{Y}_{1,i} - g(Y_{2,i}, \theta))^2
$$

(32)

where $\hat{Y}_{1,i} = n_1^{-1} P_{k_1}^i(Z_i) Q_{n_1, k_1} \sum_{i \in I_1} Y_{1,i} P_{k_1}(Z_i)$ for any $i \in I_2$ is the predicted value of $Y_{1,i}$ in the second data set based on a nonparametric regression using the first data set. Since $v_i^*$ is independent of $(X_{1,i}^*, X_{2,i}^*)$ by the data generating mechanism, $E[v_i|Y_{2,i}] = 0$ for any $i$. The nonlinear regression estimators of $\theta_0$ is consistent if $(Y_{1,i}, Y_{2,i})$ is jointly observed. Therefore, $\hat{\theta}_{Y_2,n}$ and $\hat{\theta}_{Y_1,n}$ should be consistent if $\hat{Y}_{2,i}$ and $\hat{Y}_{1,i}$ are replaced by their "true values" $Y_{2,i}$ and $Y_{1,i}$ respectively. In the simulation studies, we investigate the performances of the nonlinear regression estimators $\hat{\theta}_{Y_2,n}$ and $\hat{\theta}_{Y_1,n}$ when the missing variables are imputed. In the SMD estimation and the (imputed) nonlinear regression, we set $k_j = \text{Round}(n_j^{1/5})$ and $P_{k_j}(z) = (1, z, \ldots, z^{k_j})$. The minimization problem in the SMD estimation and the nonlinear regressions (in (31) and (32)) are solved by a grid search with $\Theta = [0, 2]$ and equally spaced grid points with grid length 0.001.

Figure 3: Properties of the Confidence Intervals ($n_1 = n_2$)

In Figures 1 and 2 we present the finite sample properties of the identity weighted SMD estimator (the green dashed line), the optimal weighted SMD estimator (the black solid line), the $Y_2$-imputed estimator (the blue dotted line) and the $Y_1$-imputed estimator (the red dash-dotted line). In Figure 1 we see that the bias and variance of the two SMD estimators converge to zero with the growth
of both $n_1$ and $n_2$. The optimal weighted SMD estimator has a slightly larger bias and much smaller variance, and much smaller root mean squared error (R-MSE) than the identity weighted SMD estimator. The improvement of the optimal SMD estimator over the identity weighted SMD estimator is clearly investigated in both model 1 and model 2. The $Y_2$-imputed estimator has almost the same finite sample bias and finite sample variance as the identity weighted SMD estimator in the linear model (i.e., model 1). But it has large and non-convergent finite sample bias in model 2, which indicates that the $Y_2$-imputed estimator may be inconsistent in general nonlinear models. The $Y_1$-imputed estimator has a large and non-convergent finite sample bias in both model 1 and model 2, which shows that it may be an inconsistent estimator in general.

Figure 4: Properties of the Confidence Intervals ($n_1 + n_2 = 1000$)

The finite sample performances of the SMD estimators and the two imputed estimators under unequal sampling schemes are presented in Figure 2. In this figure, we see that when $n_1$ is small, the finite sample bias and variance of the SMD estimators are large regardless how big $n_2$ is. This means that the main part in the estimation error of the SMD estimator is from the component estimated by the smaller sample, which is implied by Theorem 2. On the other hand, the SMD estimators are less effected when the sample size $n_2$ is small, which is because the (finite sample) variance from estimating $\phi_1(z, \theta_0)$ is larger than the finite sample variance from estimating $\phi_2(z, \theta_0)$.
in this simulation design. By (28) and the data generating mechanism

\[ u_1 \equiv Y_1 - \phi_1(Z, \theta_0) = g_2(Y_2, \theta_0) - E[g_2(Y_2, \theta_0)|Z] + v = u_2 + v \]

where \( u_2 \equiv g_2(Y_2, \theta_0) - E[g_2(Y_2, \theta_0)|Z] \) and \( E[u_2v|Z] = 0 \), which implies that

\[ \Sigma_1(z) = \Sigma_2(z) + E[v^2|Z = z] \]

and hence the conditional variance of \( u_1 \) given \( Z \) is strictly larger than the conditional variance of \( u_2 \) given \( Z \).

The finite sample properties of the inference procedures based on the identity weighted SMD estimator and the optimal weighted SMD estimator are provided in Figures 3 and 4. In Figure 3, we see that the finite coverage probabilities of the confidence intervals based on the SMD estimators converge to the nominal level 0.9 and with \( n_1 = n_2 \) starting at 50 and increasing to 1,000, and they are almost identical when \( n_1 \) and \( n_2 \) are larger than 250. In both model 1 and model 2, the coverage probabilities of the confidence intervals based on the optimal SMD estimator and the identity weighted SMD estimator are close to each other, although the confidence interval based on the optimal SMD estimator has slightly larger coverage probabilities when \( n_1 \) and \( n_2 \) are relatively small. In both models, the average length of the confidence interval of the optimal SMD estimator is much smaller than that of the confidence interval of the identity weighted SMD estimator, which is because the optimal SMD estimator has smaller variance. The finite sample performances of the confidence intervals based on the SMD estimators under unequal sampling scheme are presented in Figure 4. In this figure, we see that when \( n_1 \) (or \( n_2 \)) is small, the coverage probabilities of the confidence intervals of the two SMD estimators are far from the nominal level. The (coverage) performance of the inference based on the optimal weighted SMD estimator is slightly worse when the sample size \( n_2 \) is small regardless of the size of the other sample \( n_1 \). Figure 4 shows that the average length of the confidence intervals of the optimally weighted SMD estimators is smaller than the identity weighted SMD estimator.

7 Conclusion

This paper studies estimation and inference of nonlinear econometric models when the economic variables of the models are contained in different data sets in practice. We provide a semiparametric minimum distance estimator based on conditional moment restrictions with common conditioning variables which are contained in different data sets. The proposed estimator is shown to be consistent and has asymptotic normal distribution. We provide the specific form of optimal weight for the semiparametric minimum distance estimation, and show that the optimal weighted minimum distance estimator has the smallest asymptotic variance among all minimum distance estimators.
A consistent estimator of the variance-covariance matrix of the minimum distance estimator, and hence inference procedure of the unknown parameter is also provided. The finite sample performances of the minimum distance estimators and the inference procedure are investigated in simulation studies. The results indicate that our proposed estimator performs very well even when the sample sizes \((n_1, n_2)\) of the two data sets are quite small.

References


Hurst, E. and F. Stafford (2004). Home is where the equity is: Mortgage refinancing and household consumption. *Journal of Money, Credit, and Banking* 36(6), 985–1014.


APPENDIX

A. Proof of the Main Results in Section 3

Proof of Theorem 1. By the triangle inequality, the Cauchy-Schwarz inequality and Assumption [iv],
\[
\sup_{z \in Z} \lambda_{\max}(\hat{W}_n(z)) \leq \sup_{z \in Z} \left\| \hat{W}_n(z) - W_n(z) \right\| + \sup_{z \in Z} \lambda_{\max}(W_n(z)) \leq 2C
\]  
(33)
with probability approaching 1. By Assumption [i] and Jensen’s inequality for \( j = 1, 2, \)
\[
\sup_{\theta \in \Theta} E \left[ \|\phi_j(Z, \theta)\|^2 \right] = \sup_{\theta \in \Theta} E \left[ \|E [g_j(Y_j, \theta) \mid Z]\|^2 \right] \leq \sup_{\theta \in \Theta} E \left[ \|g_j(Y_j, \theta)\|^2 \right] \leq C,
\]
(34)
which together with Assumption [i] implies that
\[
\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \|\phi_j(Z_i, \theta)\|^2 = O_p(1).
\]
(35)
For any \( \theta \in \Theta, \) define
\[
\hat{L}_n(\theta) \equiv n^{-1} \sum_{i \in I} \left[ \hat{\phi}(Z_i, \theta)' \hat{W}_n(Z_i) \hat{\phi}(Z_i, \theta) \right]
\]
(36)
and
\[
L_n(\theta) = n^{-1} \sum_{i \in I} \phi(Z_i, \theta)' \hat{W}_n(Z_i) \phi(Z_i, \theta).
\]
(37)
By definition,
\[
\hat{L}_n(\theta) - L_n(\theta) = n^{-1} \sum_{i \in I} \left( \hat{\phi}_1(Z_i, \theta) - \phi_1(Z_i, \theta) \right)' \hat{W}_n(Z_i) \left( \hat{\phi}_1(Z_i, \theta) - \phi_1(Z_i, \theta) \right) + n^{-1} \sum_{i \in I} \left( \hat{\phi}_2(Z_i, \theta) - \phi_2(Z_i, \theta) \right)' \hat{W}_n(Z_i) \left( \hat{\phi}_2(Z_i, \theta) - \phi_2(Z_i, \theta) \right)
- 2n^{-1} \sum_{i \in I} \left( \hat{\phi}_1(Z_i, \theta) - \phi_1(Z_i, \theta) \right)' \hat{W}_n(Z_i) \left( \hat{\phi}_2(Z_i, \theta) - \phi_2(Z_i, \theta) \right)
- 2n^{-1} \sum_{i \in I} \left( \hat{\phi}_2(Z_i, \theta) - \phi_2(Z_i, \theta) \right)' \hat{W}_n(Z_i) \phi(Z_i, \theta) + 2n^{-1} \sum_{i \in I} \left( \hat{\phi}_1(Z_i, \theta) - \phi_1(Z_i, \theta) \right)' \hat{W}_n(Z_i) \phi(Z_i, \theta).
\]
(38)
Therefore, by the triangle inequality and the Cauchy-Schwarz inequality, Assumption [ii], (33), (35) and (38), we get
\[
\sup_{\theta \in \Theta} \left| \hat{L}_n(\theta) - L_n(\theta) \right| = o_p(1).
\]
(39)
By the triangle inequality, the Cauchy-Schwarz inequality, Assumption \([4](iv)\) and \([55]\),

\[
\sup_{\theta \in \Theta} |L_n(\theta) - L^*_n(\theta)| = \sup_{\theta \in \Theta} \left| n^{-1} \sum_{i \in I} \phi(Z_i, \theta)(\widetilde{W}_n(Z_i) - W_n(Z_i))\phi(Z_i, \theta) \right|
\leq \sup_{z \in \mathcal{Z}} |\widetilde{W}_n(z) - W_n(z)| \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \|\phi(Z_i, \theta)\|^2 = o_p(1), \tag{40}
\]

where \(L^*_n(\theta) = n^{-1} \sum_{i \in I} \phi(Z_i, \theta)W_n(Z_i)\phi(Z_i, \theta)\). For any \(\varepsilon > 0\), by the definition of \(\hat{\theta}_n\),

\[
\Pr \left( \|\hat{\theta}_n - \theta_0\| \geq \varepsilon \right) 
\leq \Pr \left( \inf_{\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}} \hat{L}_n(\theta) \leq \hat{L}_n(\theta_0) \right) 
\leq \Pr \left( \inf_{\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}} L_n(\theta) \leq L_n(\theta_0) + 2 \sup_{\theta \in \Theta} \left| \hat{L}_n(\theta) - L_n(\theta) \right| \right) 
\leq \Pr \left( \inf_{\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}} L^*_n(\theta) \leq L^*_n(\theta_0) + 2 \sup_{\theta \in \Theta} \left| L_n(\theta) - L^*_n(\theta) \right| + o_p(1) \right) 
\leq \Pr \left( \inf_{\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}} n^{-1} \sum_{i \in I} \|\phi(Z_i, \theta)\|^2 \leq \frac{o_p(1)}{\inf_{z \in \mathcal{Z}} \lambda_{\min}(W_n(z))} \right) 
\leq \Pr \left( \inf_{\{\theta \in \Theta : \|\theta - \theta_0\| \geq \varepsilon\}} E[\|\phi(Z, \theta)\|^2] \leq \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \|\phi(Z_i, \theta)\|^2 - E[\|\phi(Z, \theta)\|^2] + o_p(1) \right) 
\leq \Pr (\eta \leq o_p(1)) \tag{41}
\]

where the first inequality is by the definition of \(\hat{\theta}_n\), the second inequality is by

\[- \sup_{\theta \in \Theta} \left| \hat{L}_n(\theta) - L_n(\theta) \right| \leq \hat{L}_n(\theta) - L_n(\theta) \leq \sup_{\theta \in \Theta} \left| \hat{L}_n(\theta) - L_n(\theta) \right| \]

for any \(\theta \in \Theta\), the third inequality is by \([39]\), the fourth inequality is by \([40]\), the fifth inequality is by Assumption \([1](iv)\) and the last inequality is by Assumption \([1](v)\). Since \(\eta\) is a fixed positive constant, by \([41]\) we deduce that \(\Pr \left( \|\hat{\theta}_n - \theta_0\| \geq \varepsilon \right) \to 0 \) as \(n_1, n_2 \to \infty\), which proves the consistency of \(\hat{\theta}_n\). \(\blacksquare\)

For \(j = 1, 2\) and for \(m = 1, \ldots, d_g\), let \(\hat{\phi}_{j,m}(z, \theta)\) denote the \(m\)-th component in \(\hat{\phi}_j(z, \theta)\). Define

\[
\hat{\phi}_{j,m}(z, \theta) = \frac{\partial^2 \hat{\phi}_{j,m}(z, \theta)}{\partial \theta \partial \theta'} \text{ for any } \theta \in \Theta.
\]

By definition, \(\hat{\phi}_{j,m}^{(2)}(z, \theta) = P_{k_j}(z)/(P'_{nj,k_j} P_{nj,k_j})^{-1} \sum_{i \in I_j} P_{k_j}(Z_i)g_m^{(2)}(Y_{ji}, \theta)\).

**Lemma 3.** By Assumptions \([2](i), [2](ii), [2](iv)\) and \([3](iii)\), we have

\[
\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j,m}^{(2)}(Z_i, \theta) \right\|^2 = O_p(1)
\]

25
for $j = 1, 2$ and $m = 1, \ldots, d_g$.

Proof of Lemma 3. By Assumption 1(i), one can use Rudelson’s law of large numbers for matrices (see, e.g., Lemma 6.2 in Belloni, et al., 2015) to show that

$$Q_{n,k_j} - Q_{k_j} = O_p(\xi_{k_j} (\log(k_j))^{1/2} n^{-1/2})$$

and $Q_{n,j} - Q_{k_j} = O_p(\xi_{k_j} (\log(k_j))^{1/2} n_j^{-1/2})$ (42)

where $Q_{n,k_j} = n^{-1} \sum_{i \in I} P_k(Z_i) P_{k_j}(Z_i)$, $Q_{n,j} = n^{-1} \sum_{i \in I_j} P_k(Z_i) P_{k_j}(Z_i)$ and the convergence is under the matrix operator norm. By (42), Assumptions 2(iv) and 3(iii),

$$(2C)^{-1} \leq \lambda_{\min}(Q_{n,k_j}) \leq \lambda_{\max}(Q_{n,k_j}) \leq 2C$$

and $$(2C)^{-1} \leq \lambda_{\min}(Q_{n,j}) \leq \lambda_{\max}(Q_{n,j}) \leq 2C,$$

with probability approaching 1. By definition,

$$\hat{\phi}^{(2)}_{j,m}(z, \theta) = n_j^{-1} P_k(z) Q_{n,j,k_j}^{-1} \sum_{i \in I_j} P_k(Z_i) g^{(2)}_m(Y_{j,i}, \theta).$$

Without loss of generality, we assume that $\theta$ is a scalar. Since $d_\theta$ is finite, the proof in the scalar case can be applied component by component to show the lemma in the case that $\theta$ is a vector. For any $\theta \in \Theta$

$$n^{-1} \sum_{i_1 \in I} ||\hat{\phi}^{(2)}_{j,m}(Z_{i_1}, \theta)||^2$$

$$= n^{-1} \sum_{i_1 \in I} ||P_k(Z_{i_1}) Q_{n,j,k_j}^{-1} \sum_{i \in I_j} P_k(Z_i) g^{(2)}_m(Y_{j,i}, \theta)||^2$$

$$= \sum_{i_1 \in I_j} g^{(2)}_m(Y_{j,i}, \theta) P_k(Z_i) (Q_{n,j,k_j} Q_{n,j,k_j})^{-1} \sum_{i \in I_j} P_k(Z_i) g^{(2)}_m(Y_{j,i}, \theta)$$

$$\leq \lambda_{\max}(Q_{n,k_j}) n_j^{-1} \sum_{i_1 \in I_j} g^{(2)}_m(Y_{j,i}, \theta) P_k(Z_{i_1}) (P'_{n,j,k_j} P_{n,j,k_j})^{-1} \sum_{i_2 \in I_j} P_k(Z_{i_2}) g^{(2)}_m(Y_{j,i_2}, \theta)$$

$$\leq \lambda_{\max}(Q_{n,k_j}) \lambda_{\min}(Q_{n,j,k_j})^{-1} \sum_{i_1 \in I_j} ||g^{(2)}_m(Y_{j,i}, \theta)||^2$$ (44)

where the second inequality is by the fact that $P_{n,j,k_j} (P'_{n,j,k_j} P_{n,j,k_j})^{-1} P'_{n,j,k_j}$ is an idempotent matrix. By Assumption 2(i), (43) and (44),

$$\sup_{\theta \in \Theta} \sum_{i \in I} n^{-1} ||\hat{\phi}^{(2)}_{j,m}(Z_{i}, \theta)||^2 = O_p(1)$$ (45)

which finishes the proof. □
Lemma 4. Under Assumptions 1(i), 2(iv), 2(v) and 3(iii), we have for $j = 1, 2$

$$C^{-1} \leq \lambda_{\min}(Q_{n_j,u_j}) \leq \lambda_{\max}(Q_{n_j,u_j}) \leq C$$

with probability approaching 1.

Proof of Lemma 4. For any $a \in R^{d_g k_j}$ with $a'a = 1$, we can write $a = (a'_1, \ldots, a'_d)^\top$ where $a_m$ ($m = 1, \ldots, d_g$) is a $k_j \times 1$ real vector. Since $a'a = 1$,

$$\sum_{m=1}^{d_g} a'_m a_m = 1, \quad (46)$$

which implies that

$$a'_m a_m \leq 1 \text{ for any } m = 1, \ldots, d_g. \quad (47)$$

Let $\Sigma_{j,m_1,m_2}(z)$ denote the component in the $m_1$-th row and $m_2$-th column of $\Sigma_j(z)$. By the triangle inequality and Cauchy-Schwarz inequality,

$$a'Q_{n_j,u_j}a = \sum_{1 \leq m_1, m_2 \leq d_g} n_j^{-1} \sum_{i \in I_j} \left| \Sigma_{j,m_1,m_2}(z_i) a'_m (P_{k_j}(Z_i)P'_{k_j}(Z_i)) a_m \right| \leq \max_{1 \leq m_1, m_2 \leq d_g} \sup_{z \in Z} \left| \Sigma_{j,m_1,m_2}(z) \right| \sum_{1 \leq m_1, m_2 \leq d_g} n_j^{-1} \sum_{i \in I_j} \left| a'_m (P_{k_j}(Z_i)P'_{k_j}(Z_i)) a_m \right|$$

$$\leq \sup_{z \in Z} tr(\Sigma_j(z)) \sum_{1 \leq m_1, m_2 \leq d_g} \left( n_j^{-1} \sum_{i \in I_j} \left| a'_m P_{k_j}(Z_i) \right|^2 n_j^{-1} \sum_{i \in I_j} \left| a'_m P_{k_j}(Z_i) \right|^2 \right)^{1/2}$$

$$= \sup_{z \in Z} tr(\Sigma_j(z)) \sum_{1 \leq m_1, m_2 \leq d_g} \left( (a'_{m_1} Q_{n_j,k_j} a_{m_1})(a'_{m_2} Q_{n_j,k_j} a_{m_2}) \right)^{1/2} \leq d_g^2 \sup_{z \in Z} tr(\Sigma_j(z)) \lambda_{\max}(Q_{n_j,k_j}) \leq C d_g^2 \quad (48)$$

with probability approaching 1, where the fourth inequality is by (47) and the fifth inequality is by Assumption 2(v) and (43) in the proof of Lemma 3. By Assumptions 2(iv) and 2(v), there exist matrices $D_{1,j}(z)$ and $D_{2,j}$ with $D_{1,j}(z)D_{1,j}(z)' = I_{d_g}$ and $D_{2,j}D_{2,j}' = I_{k_j}$ such that

$$D_{1,j}(z)\Sigma_j(z)D_{1,j}(z)' = \Lambda_j(z) \text{ and } D_{2,j}Q_{k_j}D_{2,j}' = \Lambda Q_j \quad (49)$$

where $\Lambda_j(z)$ and $\Lambda Q_j$ are diagonal matrices with eigenvalues of $\Sigma_j(z)$ and $Q_{k_j}$ on their diagonals respectively. By definition,

$$(D_{1,j}(z) \otimes D_{2,j}) \times (D_{1,j}(z) \otimes D_{2,j})' = I_{d_g k_j} \quad (50)$$
which implies that for any \( a \in \mathbb{R}^{d_y k_j} \), there exists a (unique) \( b(z) \in \mathbb{R}^{d_y k_j} \) such that

\[
a = (D_{1,j}(z) \otimes D_{2,j}')b(z).
\]

Therefore,

\[
a^\prime \left( n_j^{-1} \sum_{i \in I_j} \Sigma_j(Z_i) \otimes (P_{k_j}(Z_i) P_{k_j}'(Z_i)) \right) a \\
= n_j^{-1} \sum_{i \in I_j} a^\prime \left( \Sigma_j(Z_i) \otimes (P_{k_j}(Z_i) P_{k_j}'(Z_i)) \right) a \\
= n_j^{-1} \sum_{i \in I_j} b(Z_i)^\prime \left( \Lambda_j(Z_i) \otimes (D_{2,j} P_{k_j}(Z_i) P_{k_j}'(Z_i) D_{2,j}') \right) b(Z_i) \\
\geq \inf_{z \in \mathbb{Z}} \lambda_{\min}(\Sigma_j(z)) n_j^{-1} \sum_{i \in I_j} b(Z_i)^\prime \left( I_m \otimes (D_{2,j} P_{k_j}(Z_i) P_{k_j}'(Z_i) D_{2,j}') \right) b(Z_i).
\]

By (50) and (51), \( b(z) = (D_{1,j}(z) \otimes D_{2,j})a \) which implies that

\[
n_j^{-1} \sum_{i \in I_j} b(Z_i)^\prime \left( I_m \otimes (D_{2,j} P_{k_j}(Z_i) P_{k_j}'(Z_i) D_{2,j}') \right) b(Z_i) \\
= n_j^{-1} \sum_{i \in I_j} a^\prime(D_{1,j}(Z_i) \otimes D_{2,j})^\prime \left( I_m \otimes (D_{2,j} P_{k_j}(Z_i) P_{k_j}'(Z_i) D_{2,j}') \right) (D_{1,j}(Z_i) \otimes D_{2,j})a \\
= a^\prime (I_m \otimes Q_{n_j,k_j}) a \geq \lambda_{\min}(Q_{n_j,k_j})
\]

which together with Assumption 2(v), (43) in the proof of Lemma 3 and (52) implies that

\[
a^\prime \left( n_j^{-1} \sum_{i \in I_j} \Sigma_j(Z_i) \otimes (P_{k_j}(Z_i) P_{k_j}'(Z_i)) \right) a \geq C^{-1}
\]

with probability approaching 1. The claim of the lemma follows from (48) and (54).  

**Lemma 5.** Under Assumptions 1(i), 1(iv), 2(iii), 2(iv), 3(v) and 3(iii), we have for \( j = 1, 2 \)

\[
(nm_j)^{-1} \phi_{w\theta,n}(I_{k_j} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1}) \sum_{i \in I_j} (u_{j,i} \otimes P_{k_j}(Z_i)) = O_p(n_j^{-1/2}).
\]

**Proof of Lemma 5** Let \( E[\cdot | \{ Z_i \}_{i \in I}] \) denote the conditional expectation given \( \{ Z_i \}_{i \in I} \). Since \( \phi_{w\theta,n} \),
By Assumption 1(i) we deduce that

$$E \left[ \left\| (n_j)^{-1} \phi_{w\theta,n}(I_d \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) \sum_{i \in I_j} (u_{j,i} \otimes P_{k_j}(Z_i)) \right\|^2 \right] \{Z_i\}_{i \in I}$$

$$= n^{-2} n_j^{-1} \text{tr} \left( \phi_{w\theta,n}(I_d \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) Q_{n,j,u_j} (I_d \otimes P_{n,k_j} Q_{n,j,k_j}^{-1})' \phi_{w\theta,n}' \right)$$

$$\leq \frac{\lambda_{\max}(Q_{n,j,u_j})}{n^2 n_j (\lambda_{\min}(Q_{n,j,k_j}))^2} \text{tr} \left( \phi_{w\theta,n}(I_d \otimes P_{n,k_j} P_{n,k_j}' P_{n,k_j}^{-1} P_{n,k_j}) \phi_{w\theta,n}' \right)$$

$$\leq \frac{\lambda_{\max}(Q_{n,j,u_j}) \lambda_{\max}(Q_{n,j,k_j})}{mn_j (\lambda_{\min}(Q_{n,j,k_j}))^2} \text{tr} \left( \phi_{w\theta,n}(I_d \otimes P_{n,k_j} P_{n,k_j}' P_{n,k_j}^{-1} P_{n,k_j}) \phi_{w\theta,n}' \right)$$

$$\leq \frac{\lambda_{\max}(Q_{n,j,u_j}) \lambda_{\max}(Q_{n,j,k_j})}{n_j (\lambda_{\min}(Q_{n,j,k_j}))^2} \text{tr} \left( n^{-1} \phi_{w\theta,n} \phi_{w\theta,n}' \right),$$

(55)

where the last inequality is by the fact that $P_{n,k_j} (P_{n,k_j}' P_{n,k_j}^{-1} P_{n,k_j})$ is an idempotent matrix. By the definition of $\phi_{w\theta,n}$,

$$n^{-1} \phi_{w\theta,n} \phi_{w\theta,n}' = n^{-1} \sum_{m=1}^{d_d} \phi_{w\theta,m,n} \phi_{w\theta,m,n}'$$

$$= n^{-1} \sum_{m=1}^{d_d} \sum_{i \in I} \phi^{(1)}(Z_i,\theta_0)' W_{n,m}(Z_i) W_{n,m}(Z_i)' \phi^{(1)}(Z_i,\theta_0)$$

$$= n^{-1} \sum_{i \in I} \phi^{(1)}(Z_i,\theta_0)' W_n(Z_i) \phi^{(1)}(Z_i,\theta_0)$$

(56)

which implies that

$$\text{tr} \left( n^{-1} \phi_{w\theta,n} \phi_{w\theta,n}' \right) \leq \sup_{x \in \mathcal{Z}} \lambda_{\max}(W_n(x)) \text{tr} \left( n^{-1} \sum_{i \in I} \phi^{(1)}(Z_i,\theta_0)' \phi^{(1)}(Z_i,\theta_0) \right).$$

(57)

By Assumption 2(iii) and Jensen’s inequality,

$$E \left[ \left\| \partial \phi_j(Z,\theta_0) / \partial \theta_l \right\|^2 \right] = E \left[ \left\| E \left[ \partial g_j(Y_j,\theta_0) / \partial \theta_l | Z \right] \right\|^2 \right] \leq E \left[ \left\| \partial g_j(Y_j,\theta_0) / \partial \theta_l \right\|^2 \right] \leq C,$$

(58)

which together with Assumption 1(i) and the Markov inequality implies that

$$n^{-1} \sum_{i \in I} \left\| \partial \phi_j(Z_i,\theta_0) / \partial \theta_l \right\|^2 = O_p(1).$$

(59)

By (59), the triangle inequality and the Cauchy-Schwarz inequality,

$$\text{tr} \left( n^{-1} \sum_{i \in I} \phi^{(1)}(Z_i,\theta_0)' \phi^{(1)}(Z_i,\theta_0) \right) = 2 \sum_{j=1,2} \sum_{l=1}^{d_d} n^{-1} \sum_{i \in I} \left\| \partial \phi_j(Z_i,\theta_0) / \partial \theta_l \right\|^2 = O_p(1)$$

(60)
which combined with Assumption 1(iv) and (57) implies that

$$tr \left( n^{-1} \phi_{w \theta,n} \phi'_{w \theta,n} \right) = O_p(1).$$

(61)

The claim of the lemma follows from (43), (55), Lemma 4 and the Markov inequality.

**Lemma 6.** Under Assumptions 1(i), 2(iv), 2(v), 3(ii) and 3(iii), we have for $j = 1, 2$,

$$n^{-1} \sum_{i \in I} \left\| \hat{\phi}_j(Z_i, \theta_0) - \phi_j(Z_i, \theta_0) \right\|^2 = O_p(k_j n_j^{-1} + k_j^{-2r_j}).$$

Proof of Lemma 6. The proof follows the standard arguments of showing the convergence rates of the series nonparametric regression estimators (see, e.g., Andrews (1991) and Newey (1997)). We include the proof here for completeness. Consider any $m = 1, \ldots, d_0$. By definition,

$$\hat{\phi}_{j,m}(Z_i, \theta_0) = P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} g_{j,m,n}(\theta_0)$$

where $g_{j,m,n}(\theta_0) = ((g_{j,m}(Z_i, \theta_0))_{i \in I_j})'$. Therefore

$$\hat{\phi}_{j,m}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0) = P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} U_{j,m,n}$$

$$+ P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \phi_{j,m,n}(\theta_0) - P_{k_j}(Z_i)' \beta_{j,m,k_j}$$

$$+ P_{k_j}(Z_i)' \beta_{j,m,k_j} - \phi_{j,m}(Z_i, \theta_0)$$

(62)

where $\phi_{j,m,n}(\theta_0) = ((\phi_{j,m}(Z_i, \theta_0))_{i \in I_j})'$, $U_{j,m,n} = (u_{j,m,1}, \ldots, u_{j,m,n})'$ and $u_{j,m,i}$ denotes the $m$-th component of $u_{j,i}$. By definition,

$$n^{-1} \sum_{i \in I} \left| P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} U_{j,m,n} \right|^2$$

$$= U_{j,m,n}' P_{n_j,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} Q_{n,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} U_{j,m,n}$$

$$= tr \left( Q_{n,k_j}^{1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} U_{j,m,n} U_{j,m,n} Q_{n,k_j}^{1/2} \right).$$

(63)

By Assumptions 1(i) and 2(v), and (43) in the proof of Lemma 3

$$tr \left( Q_{n,k_j}^{1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} E \left[ U_{j,m,n} U_{j,m,n}' \right] \{Z_i\}_{i \in I} P_{n_j,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} Q_{n,k_j}^{1/2} \right)$$

$$\leq C tr \left( Q_{n,k_j}^{1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} Q_{n,k_j}^{1/2} \right) = C n_j^{-1} tr \left( Q_{n,k_j}^{-1} Q_{n,k_j} \right) = O_p(k_j n_j^{-1})$$

(64)

which together with (63) and the Markov inequality implies that

$$n^{-1} \sum_{i \in I} \left| P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} U_{j,m,n} \right|^2 = O_p(k_j n_j^{-1}).$$

(65)
By Assumptions 1(i), 3(ii) and the Markov inequality,

\[ n^{-1} \sum_{i \in I_j} (\phi_{i,m}(Z_i, \theta_0) - P_k(Z_i)\beta_{\phi,j,m,k_j})^2 = O_p(k_j^{-2r\phi}). \] (66)

By definition,

\[ P_k(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1}P'_{n_j,k_j} \phi_{i,m,n_j}(\theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} = P_k(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_k(Z_i) \left( \phi_{i,m}(Z_i, \theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} \right). \] (67)

Therefore, by (43) in the proof of Lemma 3 and (66),

\[ n^{-1} \sum_{i \in I} \left| P_k(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1}P'_{n_j,k_j} \phi_{i,m,n_j}(\theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} \right|^2 \]
\[ = \sum_{i \in I_j} \left( \phi_{i,m}(Z_i, \theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} \right) P_k(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1}Q_{n,k_j} \]
\[ \times (P'_{n_j,k_j} P_{n_j,k_j})^{-1} \sum_{i \in I_j} P_k(Z_i) \left( \phi_{i,m}(Z_i, \theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} \right) \]
\[ \leq \frac{\lambda_{\text{max}}(Q_{n,k_j})}{n_j \lambda_{\text{min}}(Q_{n,k_j})} \sum_{i \in I_j} \left( \phi_{i,m}(Z_i, \theta_0) - P_k(Z_i)'\beta_{\phi,j,m,k_j} \right)^2 = O_p(k_j^{-2r\phi}). \] (68)

Combining the results in (62), (65), (66) and (68), we deduce that

\[ n^{-1} \sum_{i \in I} \left\| \phi_j(Z_i, \theta_0) - \phi_j(Z_i, \theta_0) \right\|^2 \]
\[ = \sum_{m=1}^{d_y} n^{-1} \sum_{i \in I} \left\| \phi_{i,m}(Z_i, \theta_0) - \phi_{i,m}(Z_i, \theta_0) \right\|^2 = O_p(k_j n_j^{-1} + k_j^{-2r\phi}), \] (69)

which finishes the proof.

For \( j = 1, 2 \), define

\[ \hat{\phi}_j^{(1)}(z, \theta) \equiv \partial \hat{\phi}_j(z, \theta)/\partial \theta \] for any \( \theta \in \Theta \) and any \( z \in Z \).

By definition, \( \hat{\phi}_j^{(1)}(z, \theta) = P_k(z)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} \sum_{i \in I_j} P_k(Z_i) g^{(1)}_j(Y_{j,i}, \theta) \).

**Lemma 7.** Under Assumptions 1(i), 3(iii), 3(iv), 3(i) and 3(iii), we have for \( j = 1, 2 \)

\[ n^{-1} \sum_{i \in I} \left\| \phi_j^{(1)}(Z_i, \theta_0) - \phi_j^{(1)}(Z_i, \theta_0) \right\|^2 = o_p(1). \]

**Proof of Lemma 7** The proof follows the standard arguments of showing the convergence rates of

31
the series nonparametric regression estimators (see, e.g., Andrews (1991) and Newey (1997)). We include the proof here for completeness. Without loss of generality, we assume that \( \theta_0 \) is a scalar, since the proof in the scalar case can be applied component by component to show the claim in the case that \( \theta_0 \) is a vector. Consider any \( m = 1, \ldots, d_g \). By Assumption 3(i) and (58), there exists \( \beta_{\phi_{g}}, m, k_j \in R^{k_j} \), such that as \( k_j \to \infty \),

\[
E \left[ \left| \phi_{j,m}(Z, \theta_0) - P_{k_j}(Z)' \beta_{\phi_{g}}, m, k_j \right|^2 \right] = o(1),
\]

which together with Assumption 1(i) and the Markov inequality implies that

\[
n^{-1} \sum_{i \in I} \left| \phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i)' \beta_{\phi_{g}}, m, k_j \right|^2 = o_p(1).
\]

By definition

\[
\hat{\phi}_{j,m}(z, \theta_0) = P_{k_j}(z)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} g_{j,m,n_j}(\theta_0)
\]

where \( g_{j,m,n_j}(\theta_0) = ((g_{j,m}(Y_{j,i}, \theta))'_{i \in I_j})' \) and \( g_{j,m}(Y_{j,i}, \theta) = \partial g_{j,m}(Y_{j,i}, \theta)/\partial \theta' \). Then

\[
\hat{\phi}_{j,m}(z, \theta_0) - \phi_{j,m}(Z_i, \theta_0) = P_{k_j}(z)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} \left( g_{j,m,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0) \right) + P_{k_j}(z)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} \left( \phi_{j,m,n_j}(\theta_0) - \phi_{j,m,k_j,n_j} \right) + (\phi_{j,m,k_j}(z) - \phi_{j,m}(z, \theta_0))'
\]

where \( \phi_{j,m,n_j}(\theta_0) = ((\phi_{j,m}(Z_i, \theta_0))'_{i \in I_j})' \),

\[
\phi_{j,m,k_j}(z) = P_{k_j}(z)' \beta_{\phi_{g}}, m, k_j \quad \text{and} \quad \phi_{j,m,k_j,n_j} = ((\phi_{j,m,k_j}(Z_i))'_{i \in I_j})'.
\]

By (43) in the proof of Lemma 3 and (71),

\[
n^{-1} \sum_{i \in I} \left| P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} \left( \phi_{j,m,n_j}(\theta_0) - \phi_{j,m,k_j,n_j} \right) \right|^2 \leq \frac{\lambda_{\max}(Q_{n,k_j})}{\lambda_{\min}(Q_{n,k_j})} \text{tr} \left( \left( P_{n_j,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} \left( \phi_{j,m,n_j}(\theta_0) - \phi_{j,m,k_j,n_j} \right) \right)^2 \right)
\]

\[
\leq \frac{\lambda_{\max}(Q_{n,k_j})}{\lambda_{\min}(Q_{n,k_j})} n_j^{-1} \sum_{i \in I_j} \left| \phi_{j,m,k_j}(Z_i) - \phi_{j,m}(Z_i, \theta_0) \right|^2 = o_p(1),
\]

where the second inequality is by the fact that \( P_{n_j,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P_{n_j,k_j} \) is an idempotent matrix.
By Assumptions 1(i), 2(iii) and (43),
\[
E \left[ n^{-1} \sum_{i \in I} \left\| P_{kj}(Z_i)'(P'_{n_{kj}} P_{n_{kj},kj})^{-1} P'_{n_{kj},kj} (g_{j,m,n_j}^{(1)}(\theta_0) - \phi_{j,m,n_j}^{(1)}(\theta_0)) \right\|^2 \right] \{ Z_i \}_{i \in I} 
\]
which together with the Markov inequality implies that
\[
\sup_{Z \in \mathcal{Z}} E \left[ \left\| g_{j,m}(Y_j, \theta_0) \right\|^2 | Z = z \right] \leq \frac{1}{n} \text{tr}(Q_{n,k_j} Q_{n,k_j}^{-1}) = O_p(k_j n_j^{-1}) \quad (74)
\]
which together with Assumption 3(i) implies that there exists \( \theta \) and \( \phi \) such that
\[
\sup_{\theta,m,n} \left( \sum_{z,\theta} (\phi_{j,m,n}(z, \theta))/m, n, m \right) \leq \mathcal{O}_p(1). \quad (75)
\]
The claim of the lemma now follows from Assumption 3(iii), (71), (72), (73) and (75).

Lemma 8. By Assumptions 1(i), 4(iv), 2(iii), 2(iv), 3(i) and 3(iii), we have for \( j = 1, 2 \) and \( m = 1, \ldots, d_g \)
\[
\sum_{m=1}^{d_g} n^{-1} \phi_{w \theta, m, n} P_{n_{kj}} (P'_{n_{kj}} P_{n_{kj},kj})^{-1} P'_{n_{kj},kj} \phi_{w \theta, m, n} = E \left[ \phi^{(1)}(Z, \theta_0) W_n(Z) \phi^{(1)}(Z, \theta_0) \right] + o_p(1). \quad (76)
\]
Proof of Lemma 8. Let \( \phi^{(1)}_{\theta_l}(z, \theta_0) = \partial \phi(z, \theta_0) / \partial \theta_l \) and \( \phi_{w \theta, m, n} = (\phi^{(1)}_{\theta_l}(Z, \theta_0)^{w \theta, m, n})_{i \in I} \) for \( l = 1, \ldots, d_\theta \). For \( m_1 = 1, \ldots, d_g \), let \( \phi_{\theta_{l,m_1}}^{(1)}(z, \theta_0) \) and \( W_{n,m_1}(z) \) denote the \( m_1 \)-th components of \( \phi_{\theta_l}^{(1)}(z, \theta_0) \) and \( W_{n,m}(z) \) respectively. By Assumption 2(iv) and (58),
\[
E \left[ W_{n,m_1}(Z) \phi_{\theta_{l,m_1}}^{(1)}(Z, \theta_0) \right]^2 \leq \sup_{z \in \mathcal{Z}} |W_{n,m_1}(z)|^2 E \left[ \phi_{\theta_{l,m_1}}^{(1)}(Z, \theta_0) \right]^2 \leq (\sup_{z \in \mathcal{Z}} (\lambda_{\max}(W_n(z))))^2 E \left[ \phi_{\theta_{l,m_1}}^{(1)}(Z, \theta_0) \right]^2 \leq C \quad (76)
\]
which together with Assumption 3(i) implies that there exists \( \beta_{l,m_1,k,j,n} \in \mathbb{R}^{d_g} \), such that
\[
E \left[ W_{n,m_1}(Z) \phi_{\theta_{l,m_1}}^{(1)}(Z, \theta_0) - P_{kj}(Z) \phi_{\theta_{l,m_1}}^{(1)}(Z, \theta_0) \beta_{l,m_1,k,j,n} \right]^2 = o(1). \quad (77)
\]
By Assumption 1(i), (77) and the Markov inequality,
\[
n^{-1} \sum_{i \in I} \left| W_{n,m_1}(Z_i) \phi_{\theta_{l,m_1}}^{(1)}(Z_i, \theta_0) - P_{kj}(Z_i) \phi_{\theta_{l,m_1}}^{(1)}(Z_i, \theta_0) \beta_{l,m_1,k,j,n} \right|^2 = o_p(1). \quad (78)
\]
By (78) and the definition of $\phi_{w\theta,m,n}$,

$$n^{-1} \phi_{w\theta_1,m,n} (I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}) \phi'_{w\theta_1,m,n} = n^{-1} \sum_{i \in I} \left| \phi^{(1)}_{\theta_i} (Z_i, \theta_0)' W_{n,m} (Z_i) - P_{k_j} (Z_i)' (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j} \phi'_{w\theta_1,m,n} \right|^2 \leq 4d_g \sum_{l=1}^{d_g} n^{-1} \sum_{i \in I} \left| W_{n,m,m_1} (Z_i) \phi^{(1)}_{\theta_{l,m_1}} (Z_i, \theta_0) - P_{k_j} (Z_i)' \beta_{l,m_1,k_j,n} \phi'_{w\theta_1,m,n} \right|^2 = o_p(1). \quad (79)$$

For any $l_1, l_2 = 1, \ldots, d_\theta$, since $I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}$ is idempotent, by Cauchy-Schwarz inequality and (79),

$$\left| n^{-1} \phi_{w\theta_1,m,n} (I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}) \phi'_{w\theta_1,m,n} \right|^2 \leq \frac{\phi_{w\theta_1,m,n} (I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}) \phi'_{w\theta_1,m,n}}{n} \times \frac{\phi_{w\theta_1,m,n} (I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}) \phi'_{w\theta_1,m,n}}{n} = o_p(1)$$

which implies that

$$n^{-1} \phi_{w\theta,m,n} (I_n - P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j}) \phi'_{w\theta,m,n} = o_p(1). \quad (80)$$

By (56) and (80),

$$\sum_{m=1}^{d_g} n^{-1} \phi_{w\theta,m,n} P_{n,k_j} (P_{n,k_j} P_{n,k_j})^{-1} P_{n,k_j} \phi'_{w\theta,m,n} = \sum_{m=1}^{d_g} n^{-1} \phi_{w\theta,m,n} \phi'_{w\theta,m,n} + o_p(1) = n^{-1} \sum_{i \in I} \phi^{(1)}_{\theta_i} (Z_i, \theta_0)' W_{n,m} (Z_i) \phi^{(1)}_{\theta_i} (Z_i, \theta_0) + o_p(1). \quad (81)$$

By Assumption 3(iii) and Jensen’s inequality,

$$E \left[ \| \partial \phi_j (Z, \theta_0) / \partial \theta_l \|^4 \right] = E \left[ \| E \left[ \partial g_j (Y_j, \theta_0) / \partial \theta_l | Z \right] \|^4 \right] \leq d_g E \left[ \| \partial g_j (Y_j, \theta_0) / \partial \theta_l \|^4 \right] \leq C, \quad (82)$$

for any $j = 1, 2$ and any $l = 1, \ldots, d_\theta$. By Cauchy-Schwarz inequality and Hölder’s inequality for
any \(l_1, l_2 = 1, \ldots, d_\theta\),

\[
E \left[ \left| \mathbf{\phi}^{(1)}_{\theta_1}(Z, \theta_0) \mathbf{W}_n(Z) \mathbf{\phi}^{(1)}_{\theta_2}(Z, \theta_0) \right|^2 \right] \\
\leq E \left[ \mathbf{\phi}^{(1)}_{\theta_1}(Z, \theta_0)' \mathbf{W}_n(Z) \mathbf{\phi}^{(1)}_{\theta_1}(Z, \theta_0) \right] \\
\leq \sup_{z \in \mathcal{Z}} (\lambda_{\max}(W_n(z)))^2 E \left[ \| \mathbf{\phi}^{(1)}_{\theta_1}(Z, \theta_0) \|^2 \| \mathbf{\phi}^{(1)}_{\theta_2}(Z, \theta_0) \|^2 \right] \\
\leq \sup_{z \in \mathcal{Z}} (\lambda_{\max}(W_n(z)))^2 E \left[ \| \mathbf{\phi}^{(1)}_{\theta_1}(Z, \theta_0) \|^4 \right] E \left[ \| \mathbf{\phi}^{(1)}_{\theta_2}(Z, \theta_0) \|^4 \right] \leq C
\]  

(83)

where the last inequality is by Assumption 1(iv) and (82). By Assumption 1(i), (83) and the Markov inequality,

\[
n^{-1} \sum_{i \in I} \mathbf{\phi}^{(1)}(Z_i, \theta_0)' \mathbf{W}_n(Z_i) \mathbf{\phi}^{(1)}(Z_i, \theta_0) = E \left[ \mathbf{\phi}^{(1)}(Z, \theta_0)' \mathbf{W}_n(Z) \mathbf{\phi}^{(1)}(Z, \theta_0) \right] + O_p(n^{-1/2}).
\]  

(84)

The claim of the lemma follows from (81) and (84).  

**Lemma 9.** Let \(\hat{\mathbf{\phi}}^{(1)}(z, \theta) = \hat{\mathbf{\phi}}^{(1)}(z, \theta) - \bar{\mathbf{\phi}}^{(1)}(z, \theta)\). Suppose \(\theta_n \in \Theta\) for all \(n\) and \(\| \theta_n - \theta_0 \| = O_p(\delta_n)\) where \(\delta_n = o(1)\) is a non-negative real sequence. Under Assumptions 1(i), 1(iv), 2(iii), 2(iv) and 3(iii), we have for \(l = 1, \ldots, d_\theta\),

\[
n^{-1} \sum_{i \in I} \hat{\mathbf{\phi}}^{(1)}_{\theta_l}(Z_i, \theta_0)' \mathbf{W}_n(Z_i) \mathbf{\phi}^{(1)}(Z_i, \theta_0) = E \left[ \mathbf{\phi}^{(1)}_{\theta_l}(Z, \theta_0)' \mathbf{W}_n(Z) \mathbf{\phi}^{(1)}(Z, \theta_0) \right] + o_p(1),
\]

where \(\hat{\mathbf{\phi}}^{(1)}_{\theta_l}(z, \theta)\) and \(\mathbf{\phi}^{(1)}_{\theta_l}(z, \theta)\) denote the \(l\)-th columns of \(\hat{\mathbf{\phi}}^{(1)}(z, \theta)\) and \(\mathbf{\phi}^{(1)}(z, \theta)\) respectively.

**Proof of Lemma 9.** We first show that

\[
n^{-1} \sum_{i \in I} \left\| \hat{\mathbf{\phi}}^{(1)}_j(Z_i, \theta_0) - \mathbf{\phi}^{(1)}_j(Z_i, \theta_0) \right\|^2 = o_p(1).
\]  

(85)

For this purpose we assume that \(\theta\) is a scalar without loss of generality, since the proof in the scalar case can be applied component by component to show (85) in the case that \(\theta\) is a vector. By definition,

\[
\hat{\mathbf{\phi}}^{(1)}_{j,m}(Z_i, \theta) = P_{k_j}(Z_i)'(P_{n_j,k_j}^{-1}P_{n_j,k_j}'g_{j,m,n}^{(1)}(\theta))\]

where \(g_{j,m,n}^{(1)}(\theta) = (\partial g_{j,m}(Y_{j,i}, \theta)/\partial \theta)'_i \in I_j\). By the first-order Taylor expansion,

\[
g_{j,m}^{(1)}(Y_{j,i}, \theta_0) - g_{j,m}^{(1)}(Y_{j,i}, \theta_0) = g_{j,m}^{(2)}(Y_{j,i}, \theta_0)(\theta_n - \theta_0)
\]  

(86)

where \(\theta_n, 0\) is between \(\theta_n\) and \(\theta_0\) and it may be different across rows. By (86) and the Cauchy-
Schwarz inequality,
\[ \left\| g_{j,m}^{(1)}(Y_{j,i}, \theta_n) - g_{j,m}^{(1)}(Y_{j,i}, \theta_0) \right\| \leq d_\theta \sup_{\theta \in \Theta} \left\| g_{j,m}^{(2)}(Y_{j,i}, \theta) \right\| \| \theta_n - \theta_0 \|. \]  
(87)

Since \( P_{n,j,k}(P'_{n,j,k} P_{n,j,k})^{-1} P'_{n,j,k} \) is an idempotent matrix,
\[ n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j,m}^{(1)}(Z_i, \theta_n) - \hat{\phi}_{j,m}^{(1)}(Z_i, \theta_0) \right\|^2 \leq \frac{\lambda_{\text{max}}(Q_{n,k})}{\lambda_{\text{min}}(Q_{n,k})} n_j^{-1} \sum_{i \in I_j} \left\| g_{j,m}^{(1)}(Y_{j,i}, \theta_n) - g_{j,m}^{(1)}(Y_{j,i}, \theta_0) \right\|^2 \]
\[ \leq \frac{\lambda_{\text{max}}(Q_{n,k}) d_\theta^2}{\lambda_{\text{min}}(Q_{n,k})} \| \theta_n - \theta_0 \|^2 \sup_{\theta \in \Theta} n_j^{-1} \sum_{i \in I_j} \left\| g_{j,m}^{(2)}(Y_{j,i}, \theta) \right\|^2 = O_p(\delta_n^2), \]  
(88)

where the second inequality is by (87), and the equality is by \( \| \theta_n - \theta_0 \| = O_p(\delta_n) \) and Assumption 2(i). By (88) and Lemma 7,
\[ n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j}^{(1)}(Z_i, \theta_n) - \phi_{j}^{(1)}(Z_i, \theta_0) \right\|^2 \leq 2 \sum_{m=1}^{d_\theta} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j,m}^{(1)}(Z_i, \theta_n) - \phi_{j,m}^{(1)}(Z_i, \theta_0) \right\|^2 \]
\[ + 2n^{-1} \sum_{m=1}^{d_\theta} \sum_{i \in I} \left\| \hat{\phi}_{j,m}^{(1)}(Z_i, \theta) - \phi_{j,m}^{(1)}(Z_i, \theta) \right\|^2 = o_p(1) \]  
(89)

which proves (85). By (59) and the definition of \( \phi_{j}^{(1)}(z, \theta_0) \),
\[ n^{-1} \sum_{i \in I} \left\| \phi_{j}^{(1)}(Z_i, \theta_0) \right\|^2 = \sum_{l=1}^{d_\theta} n^{-1} \sum_{i \in I} \| \partial \phi_{j}(Z_i, \theta_0) / \partial \theta_l \|^2 = O_p(1), \]  
(90)

which together with (85) implies that
\[ n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j}^{(1)}(Z_i, \theta_n) \right\|^2 = O_p(1). \]  
(91)

For \( j = 1, 2 \) and \( l = 1, \ldots, d_\theta \), let \( \hat{\phi}_{\theta,j}(z, \theta) \) and \( \phi_{\theta,j}(z, \theta) \) denote the \( l \)-th columns in \( \hat{\phi}_{j}^{(1)}(z, \theta) \) and \( \phi_{j}^{(1)}(z, \theta) \) respectively. For any \( j_1, j_2 = 1, 2 \) and any \( l_1, l_2 = 1, \ldots, d_\theta \), we can use the triangle
inequality to get
\[
\begin{align*}
&\left| n^{-1} \sum_{i \in I} \hat{\phi}_{t_1, j_1}(Z_i, \theta) W_n(Z_i) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \\
&- n^{-1} \sum_{i \in I} \phi_{t_1, j_1}(Z_i, \theta) W_n(Z_i) \phi_{t_2, j_2}(Z_i, \theta) \right| \\
&\leq \left| n^{-1} \sum_{i \in I} \hat{\phi}_{t_1, j_1}(Z_i, \theta) (\hat{W}_n(Z_i) - W_n(Z_i)) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right| \\
&\quad + \left| n^{-1} \sum_{i \in I} \left( \hat{\phi}_{t_1, j_1}(Z_i, \theta) - \phi_{t_1, j_1}(Z_i, \theta) \right)' W_n(Z_i) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right| \\
&\quad + \left| n^{-1} \sum_{i \in I} \phi_{t_1, j_1}(Z_i, \theta)' W_n(Z_i) \left( \hat{\phi}_{t_2, j_2}(Z_i, \theta) - \phi_{t_2, j_2}(Z_i, \theta) \right) \right|. \tag{92}
\end{align*}
\]

By the triangle inequality and the Cauchy-Schwarz inequality,
\[
\begin{align*}
&\left| n^{-1} \sum_{i \in I} \hat{\phi}_{t_1, j_1}(Z_i, \theta) (\hat{W}_n(Z_i) - W_n(Z_i)) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right| \\
&\leq \sup_{z \in Z} \| \hat{W}_n(z) - W_n(z) \| n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{t_1, j_1}(Z_i, \theta) \right\| \left\| \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right\| \\
&\leq \sup_{z \in Z} \| \hat{W}_n(z) - W_n(z) \| \max_{j=1,2} \max_{l=1,\ldots,d_\theta} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{t_1, j}(Z_i, \theta) \right\|^2 = O_p(\delta_{w,n}) \tag{93}
\end{align*}
\]

where the equality is by Assumption (iv) and (91). By the triangle inequality and the Cauchy-Schwarz inequality,
\[
\begin{align*}
&\left| n^{-1} \sum_{i \in I} \left( \hat{\phi}_{t_1, j_1}(Z_i, \theta) - \phi_{t_1, j_1}(Z_i, \theta) \right)' W_n(Z_i) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right| \\
&\leq n^{-1} \sum_{i \in I} \left\| \left( \hat{\phi}_{t_1, j_1}(Z_i, \theta) - \phi_{t_1, j_1}(Z_i, \theta) \right)' W_n(Z_i) \hat{\phi}_{t_2, j_2}(Z_i, \theta) \right\| \\
&\leq \sup_{z \in Z}(\lambda_{\max}(W_n(z))) \left( n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{t_1, j_1}(Z_i, \theta) \right\|^2 \right)^{1/2} \left\| (W_n(Z_i))^{1/2} \hat{\phi}_{t_1, j_1}(Z_i, \theta) \right\| \\
&\times \left( n^{-1} \sum_{i \in I} \left\| \phi_{t_1, j_1}(Z_i, \theta) - \phi_{t_1, j_1}(Z_i, \theta) \right\|^2 \right)^{1/2} = o_p(1) \tag{94}
\end{align*}
\]

where the equality is by Assumption (iv), (85) and (91). Similarly, we can show that
\[
\left| n^{-1} \sum_{i \in I} \phi_{t_1, j_1}(Z_i, \theta) W_n(Z_i) \left( \hat{\phi}_{t_2, j_2}(Z_i, \theta) - \phi_{t_2, j_2}(Z_i, \theta) \right)' \right| = o_p(1). \tag{95}
\]
Collecting the results in (92), (93), (94) and (95), we get

\[ n^{-1} \sum_{i \in I} \tilde{\phi}^{(1)}_{\theta_i}(Z_i, \theta_n)\tilde{W}_n(Z_i)\tilde{\phi}^{(1)}_{\theta_i}(Z_i, \theta_n) - n^{-1} \sum_{i \in I} \phi^{(1)}_{\theta_i}(Z_i, \theta_0)'W_n(Z_i)\phi^{(1)}_{\theta_i}(Z_i, \theta_0) = o_p(1) \tag{96} \]

which together with (84) proves the claim of the lemma. \[ \blacksquare \]

**Lemma 10.** Suppose \( \theta_n \in \Theta \) for all \( n \) and \( \| \theta_n - \theta_0 \| = O_p(\delta_n) \) where \( \delta_n = o(1) \) is a non-negative real sequence. Under Assumptions (i), (ii), (iii), (iv), (v), (vi), (vii) and (viii), we have

\[ n^{-1} \sum_{i \in I} \left( \phi^{(2)}_m(Z_i, \theta_n)\hat{\theta}_n - \phi^{(2)}_m(Z_i, \theta_0) \right) \hat{W}_n(Z_i)\hat{\phi}(Z_i, \theta_n) = o_p(\| \hat{\theta}_n - \theta_0 \|), \]

where \( \hat{\phi}^{(2)}_m(z, \theta) = \hat{\phi}_{1,m}(z, \theta) - \hat{\phi}_{2,m}(z, \theta) \) for \( m = 1, \ldots, d_g \).

**Proof of Lemma 10.** Let \( \hat{\phi}_{l,m}(z, \theta) \) denote the \( l \)-th row of \( \hat{\phi}^{(2)}_m(z, \theta) \) for \( l = 1, \ldots, d_g \). By the Cauchy-Schwarz inequality,

\[
\left\| n^{-1} \sum_{i \in I} \left( \phi^{(2)}_m(Z_i, \theta_n)\hat{\theta}_n - \phi^{(2)}_m(Z_i, \theta_0) \right) \hat{W}_n(Z_i)\hat{\phi}(Z_i, \theta_n) \right\|^2 \\
\leq \sup_{z \in Z}(\lambda_{\max}(\hat{W}_n(z)))^2 d_g \left\| \hat{\theta}_n - \theta_0 \right\|^2 \times n^{-1} \sum_{i \in I} \left\| \phi^{(2)}_m(Z_i, \theta_n) \right\|^2 \times n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_n) \right\|^2. \tag{97}
\]

Since \( \sup_{z \in Z}(\lambda_{\max}(\hat{W}_n(z)))^2 = O_p(1) \) by (33), to prove the lemma, it is sufficient to show that

\[ n^{-1} \sum_{i \in I} \left\| \phi^{(2)}_m(Z_i, \theta_n) \right\|^2 = O_p(1) \tag{98} \]

for \( m = 1, \ldots, d_g \), and for \( j = 1, 2 \)

\[ n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_n) \right\|^2 = o_p(1). \tag{99} \]

The claim in (98) is implied by Lemma 3 and the properties of \( \theta_n \). By the first-order expansion and the Cauchy-Schwarz inequality, there exists \( \theta_{1,n} \) between \( \theta_n \) and \( \theta_0 \) such that

\[
\left\| n^{-1} \sum_{i \in I} \left( \hat{\phi}_j(Z_i, \theta_n) - \phi_j(Z_i, \theta_0) \right)^2 \right. \\
= 2n^{-1} \sum_{i \in I} \left( \hat{\phi}_j^{(1)}(Z_i, \theta_{1,n})(\theta_n - \theta_0) \right)' \left( \hat{\phi}_j(Z_i, \theta_{1,n}) - \phi_j(Z_i, \theta_0) \right) \\
\leq 4 \left\| \theta_n - \theta_0 \right\| \left( \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_j(Z_i, \theta) \right\|^2 \times n^{-1} \sum_{i \in I} \left\| \phi^{(1)}_j(Z_i, \theta_{1,n}) \right\|^2 \right)^{1/2} = o_p(1) \tag{100}
\]

38
where the equality is by: (i) \( \| \theta_n - \theta_0 \| = o_p(1) \); (ii) \( \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \| \hat{\phi}_j(Z_i, \theta) \|^2 = O_p(1) \) which is implied by Assumption \([35](\text{ii})\) and \([35](\text{iii})\); and (iii) \( n^{-1} \sum_{i \in I} \| \hat{\phi}_j(Z_i, \theta_1, n) \|^2 = O_p(1) \) which is implied by \([35](\text{ii})\). Since \( \phi(z, \theta_0) = 0 \) for any \( z \in \mathcal{Z} \),

\[
n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_n) \right\|^2 = n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_n) - \phi(Z_i, \theta_0) \right\|^2 \\
\leq 2 \sum_{j=1,2} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_j(Z_i, \theta_n) - \phi_j(Z_i, \theta_0) \right\|^2 = o_p(1)
\]

which proves \([39](\text{i})\). 

**Lemma 11.** Under Assumptions \([7](\text{i})\), \([7](\text{iv})\), \([7](\text{iii})\), \([7](\text{iv})\), \([7](\text{v})\), \([3](\text{i})\) and \([3](\text{iii})\), we have for any \( j_1, j_2 = 1, 2 \) and \( m = 1, \ldots, d_g \),

\[
n^{-1} \sum_{i \in I} \phi_j^{(1)}(Z_i, \theta_0) W_{n,m}(Z_i) P_{k_1}(Z_i)^{\prime} (P_{n_{j_2}}^{-1} P_{n_{j_2}, k_2})^{-1} P_{k_2}(Z_i) u_{j_2,m,i} \\
= n^{-1} \sum_{i \in I} \phi_j^{(1)}(Z_i, \theta_0) W_{n,m}(Z_i) P_{k_1}(Z_i)^{\prime} (P_{n_{j_2}}^{-1} P_{n_{j_2}, k_2})^{-1} P_{k_2}(Z_i) u_{j_2,m,i} + o_p(n_1^{-1} + n_2^{-1}).
\]

**Proof of Lemma 11.** Without loss of generality, we assume that \( \theta \) is a scalar, since the proof in the scalar case can be applied component by component to show \([85](\text{iii})\) in the case that \( \theta \) is a vector. Define

\[
\Delta_{m,n,j_k}(u) \equiv (P_{n_{j_k}, k_j}^{-1} P_{n_{j_k}, k_j})^{-1} P_{k_j}(Z_i) u_{j,m,i},
\]

\[
\Upsilon_{m_1,m_2,j_k,n} \equiv n^{-1} \sum_{i \in I} W_{n,m_1,m_2}(Z_i) P_{k_j}(Z_i) P_{k_j}(Z_i)^{\prime}.
\]

Let \( E \left[ \cdot \mid \{Z_i\}_{i \in I_{j_1}}, I_{j_2} \right] \) denote the conditional expectation given \( \{Z_i\}_{i \in I_{j_1}} \) and the data set \( I_{j_2} \). By Assumptions \([1](\text{i})\), \([2](\text{v})\) and \([43]\), we have for any \( j_1 \neq j_2 \)

\[
\lambda_{\max} \left( E \left[ \Delta_{m,n,j_1,k_1}(u) \Delta_{m,n,j_1,k_1}(u)^{\prime} \mid \{Z_i\}_{i \in I_{j_1}}, I_{j_2} \right] \right) \\
\leq \sup_{z \in \mathcal{Z}} E[u_{j_1,m}^2 | Z = z] \lambda_{\max}((P_{n_{j_1}, k_1}^{-1} P_{n_{j_1}, k_1})^{-1}) \\
= \sup_{z \in \mathcal{Z}} E[u_{j_1,m}^2 | Z = z] \frac{\lambda_{\min}(Q_{j_1,k_1})}{n_{j_1}} = O_p(n_{j_1}^{-1}).
\]

(101)

We first prove the lemma in the case that \( j_1 \neq j_2 \). Without loss of generality, let \( j_1 = 1 \) and \( j_2 = 2 \).
By Assumptions 1(i), 1(iv), (43), (101) and Lemma 7,

\[
E \left[ \left| \sum_{i \in I} \phi^{(1)}_1(Z_i, \theta_0) - \phi^{(1)}_1(Z_i, \theta_0) \right| W_{n,m}(Z_i) P_{k_2}(Z_i) \Delta_{m,n_2,k_2}(u) \right|^2 \left\{ Z_i \in I_2, I_1 \right\} \right] 
\leq \sup_{z \in Z} \frac{E[u^2_{z,m} | Z = z] \lambda_{\max}(Q_{n,k_2})}{n^2 \lambda_{\min}(Q_{n_2,k_2})} \sum_{i \in I} \left( \phi^{(1)}_1(Z_i, \theta_0) - \phi^{(1)}_1(Z_i, \theta_0) \right)^2
\leq C \lambda_{\max}(Q_{n,k_2}) \sup_{z \in Z} (\lambda_{\max}(W_n(z))) \frac{1}{n^2 \lambda_{\min}(Q_{n_2,k_2})} \sum_{i \in I} \left\| \phi^{(1)}_1(Z_i, \theta_0) - \phi^{(1)}_1(Z_i, \theta_0) \right\|^2 = o_p(n^{-1}) \tag{102}
\]

which together with the Markov inequality implies that for \( j_1 \neq j_2 \),

\[
n^{-1} \sum_{i \in I} \phi^{(1)}_{j_1}(Z_i, \theta_0) W_{n,m}(Z_i) P_{k_{j_2}}(Z_i) \Delta_{m,n_{j_2},k_{j_2}}(u) = n^{-1} \sum_{i \in I} \phi^{(1)}_{j_1}(Z_i, \theta_0) W_{n,m}(Z_i) P_{k_{j_2}}(Z_i) \Delta_{m,n_{j_2},k_{j_2}}(u) + o_p(n^{-1} + n^{-1}). \tag{103}
\]

Next we prove the lemma in the case that \( j_1 = j_2 \). Without loss of generality, let \( j_1 = 1 \) and \( j_2 = 1 \). Consider any \( m_1 = 1, \ldots, d_g \). Let \( W_{n,m_1}(z) \) denote the \( m_1 \)-th component of \( W_{n,m}(z) \). For any \( a \in \mathbb{R}^{k_1} \) with \( a'a = 1 \),

\[
a'(\Upsilon_{m,m_1,k_1,n})a = n^{-1} \sum_{i \in I} W_{n,m,m_1}(Z_i)(a' P_{k_1}(Z_i))^2
\leq \sup_{z \in Z} |W_{n,m,m_1}(z)| n^{-1} \sum_{i \in I} (a' P_{k_1}(Z_i))^2
\leq \sup_{z \in Z} |\lambda_{\max}(W_n(z))| \lambda_{\max}(Q_{n,k_1}),
\]

which implies that

\[
\lambda_{\max}(\Upsilon_{m,m_1,k_1,n}) \leq \sup_{z \in Z} |\lambda_{\max}(W_n(z))| \lambda_{\max}(Q_{n,k_1}). \tag{104}
\]

Similarly, we can show that

\[
\lambda_{\min}(\Upsilon_{m,m_1,k_1,n}) \geq - \sup_{z \in Z} |\lambda_{\max}(W_n(z))| \lambda_{\max}(Q_{n,k_1}). \tag{105}
\]

By Assumption 1(iv), (43), (104) and (105),

\[
\| \Upsilon_{m,m_1,k_1,n} \|_{op} = O_p(1). \tag{106}
\]
By Assumption 2(v), (43), (70) and the Markov inequality, (101) and (106),
\[
E \sup_{n,m,n} \| \phi_{1,m}(z, \theta_0) - \phi_{1,m}(z, \theta_0) \| \leq C_k n_1 \lambda_{\min}(Q_{n,k_1}) \leq o_p(n_1^{-1}).
\]

By Assumptions 1(i), 1(iv), 2(v), (43), (71) and (101),
\[
E \left[ \left\| (g_{1,m}^{(1)}(Y_1, \theta_0) - \phi_{1,m}(Y_1, \theta_0)) P_{n,k_1} \right\|^2 \right] \leq C k_1 \lambda_{\max}(Q_{n,k_1}) = O_p(k_1),
\]

By Assumptions 1(i) and 2(iii),
\[
E \left[ \left\| (g_{1,m}^{(1)}(Y_1, \theta_0) - \phi_{1,m}(Y_1, \theta_0)) P_{n,k_1} \right\|^2 \right] \leq C k_1 \lambda_{\max}(Q_{n,k_1}) = O_p(k_1),
\]

which together with the Markov inequality implies that
\[
n_1^{-1} \left\| (g_{1,m}^{(1)}(Y_1, \theta_0) - \phi_{1,m}(Y_1, \theta_0)) P_{n,k_1} \right\|^2 = O_p(k_1).
\]

By (101),
\[
E \left[ \left\| \Delta_{n,m,k_1}(u) \right\|^2 \right] \leq k_1 \lambda_{\max} \left( E \left[ \Delta_{n,m,k_1}(u) \Delta_{n,m,k_1}(u) \right] \right) \leq k_1 \sup_{z \in \mathcal{Z}} E[u_{1,m}^2 | Z = z] / n_1 \lambda_{\min}(Q_{n,k_1})
\]

(112)
Proof of Lemma 12.

Lemma 12. Under Assumptions 1(i), 1(ii), 1(iv), 2(iii), 2(iv), 2(v), 3(ii) and 3(iii), we have

\[
\|\Delta_{m,n_1,k_1}(u)\|^2 = O_p(k_1 n_1^{-1}) .
\]  \hspace{1cm} (113)

Applying the Cauchy-Schwarz inequality and the results in (111) and (113), we get

\[
\begin{align*}
&\left\| \left(g_{1,m_1,n_1}(\theta_0) - \phi_{1,m_1,n_1}(\theta_0)\right)'P_{n_1,k_1}(P_{n_1,k_1}^{'})^{-1}\gamma_{m,m_1,k_1,n}\Delta_{m,n_1,k_1}(u) \right\|^2 \\
&\leq \frac{n_1^2}{n_1^2(\lambda_{\min}(Q_{n_1,k_1}))^2} \left\| \left(g_{1,m_1,n_1}(\theta_0) - \phi_{1,m_1,n_1}(\theta_0)\right)'P_{n_1,k_1} \right\|^2 \|\Delta_{m,n_1,k_1}(u)\|^2 \\
&= O_p(k_1^2 n_1^{-2}) = o_p(n_1^{-1}),
\end{align*}
\]  \hspace{1cm} (114)

where the second equality is by Assumption 3(iii). Collecting the results in (107), (108), (109) and (114), we have for \( j_1 = j_2 \)

\[
n^{-1} \sum_{i \in I} \phi_{j_1}(Z_i, \theta_0)'W_{n,m}(Z_i)P_{k_j}(Z_i)'\Delta_{m,n_1,k_1}(u) = n^{-1} \sum_{i \in I} \phi_{j_1}(Z_i, \theta_0)'W_{n,m}(Z_i)P_{k_j}(Z_i)'\Delta_{m,n_1,k_1}(u) + o_p(n_1^{-1} + n_2^{-1})
\]  \hspace{1cm} (115)

which finishes the proof. \( \blacksquare \)

Lemma 12. Under Assumptions 1(i), 1(ii), 1(iv), 2(iii), 2(iv), 2(v), 3(ii) and 3(iii), we have

\[
n^{-1} \sum_{i \in I} \hat{\phi}_{j_1}(Z_i, \theta_0)'\hat{W}_{n}(Z_i)\hat{\phi}(Z_i, \theta_0)
\]

\[
= (nn_1)^{-1}\phi_{w\theta,n}(I_m \otimes P_{n,k_1}Q_{n_1,k_1}^{-1}) \sum_{i \in I} u_{1,i} \otimes P_{k_1}(Z_i) - (nn_2)^{-1}\phi_{w\theta,n}(I_m \otimes P_{n,k_2}Q_{n_2,k_2}^{-1}) \sum_{i \in I} u_{2,i} \otimes P_{k_2}(Z_i) + o_p(n_1^{-1/2} + n_2^{-1/2}) .
\]

Proof of Lemma 12. By Lemma 5 and \( \phi(z, \theta_0) = 0 \) for \( z \in Z \) a.e.,

\[
n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_0) \right\|^2 = n^{-1} \sum_{i \in I} \left\| \phi(Z_i, \theta_0) - \hat{\phi}(Z_i, \theta_0) \right\|^2 \leq 2 \sum_{j=1,2} n^{-1} \sum_{i \in I} \left\| \phi_j(Z_i, \theta_0) - \phi_j(Z_i, \theta_0) \right\|^2 = O_p(max(k_1 n_1^{-1} + k_2^{-2}).
\]  \hspace{1cm} (116)

By Assumption 3(iii), Lemma 7 and (90),

\[
n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{j_1}(Z_i, \theta_0) \right\|^2 = O_p(1).
\]  \hspace{1cm} (117)

42
which together with the triangle inequality, the Cauchy-Schwarz inequality, Assumptions 1(iv) and 3(iii), (116) and (117) implies that

\[
\left\| n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \theta_0)'(\hat{W}_n(Z_i) - W_n(Z_i))\hat{\phi}(Z_i, \theta_0) \right\|^2 \\
\leq \left( \sup_{z \in Z} \left| \hat{W}_n(z) - W_n(z) \right| \right)^2 n^{-1} \sum_{i \in I} \left\| \hat{\phi}^{(1)}(Z_i, \theta_0) \right\|^2 n^{-1} \sum_{i \in I} \left\| \hat{\phi}(Z_i, \theta_0) \right\|^2 \\
= O_p(\delta_{w,n}^2 \max(k_j n_j^{-1} + k_j^{-2r_\phi})) = o_p(n_1^{-1} + n_2^{-1})
\]

which implies that

\[
n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \theta_0)'\hat{W}_n(Z_i)\hat{\phi}(Z_i, \theta_0) \\
= n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \theta_0)'W_n(Z_i)\hat{\phi}(Z_i, \theta_0) + o_p(n_1^{-1} + n_2^{-1}).
\]

By definition for \( m = 1, \ldots, d_g, \)

\[
\hat{\phi}_{j,m}(z, \theta_0) - \phi_{j,m}(z, \theta_0) = n_j^{-1} P_{k_j}(z)'Q_{n_j,k_j}^{-1} \sum_{i \in I_{j}} P_{k_j(\cdot)}(Z_i)g_{j,m}(Y_{j,i}, \theta_0) - \phi_{j,m}(z, \theta_0) \\
= n_j^{-1} P_{k_j}(z)'Q_{n_j,k_j}^{-1} \sum_{i \in I_{j}} P_{k_j(\cdot)}(Z_i)u_{j,m,i} \\
+ n_j^{-1} P_{k_j}(z)'Q_{n_j,k_j}^{-1} \sum_{i \in I_{j}} P_{k_j(\cdot)}(Z_i)(\phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i)'\beta_{\phi_j,m,k_j}) \\
+ P_{k_j}(z)'\beta_{\phi_j,m,k_j} - \phi_{j,m}(z, \theta_0).
\]

By the Cauchy-Schwarz inequality, Assumptions 1(i), 1(iv), 3(iii), (66) and (117),

\[
\left\| n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \theta_0)'W_{n,m}(Z_i)(P_{k_j(\cdot)}(Z_i)'\beta_{\phi_j,m,k_j} - \phi_{j,m}(Z_i, \theta_0)) \right\|^2 \\
\leq \left( \sup_{z \in Z} (\lambda_{\max}(W_n(z))) \right)^2 n^{-1} \sum_{i \in I} \left\| \hat{\phi}^{(1)}(Z_i, \theta_0) \right\|^2 \\
\times n^{-1} \sum_{i \in I} \left\| P_{k_j(\cdot)}(Z_i)'\beta_{\phi_j,m,k_j} - \phi_{j,m}(Z_i, \theta_0) \right\|^2 = O_p(k_j^{-2r_\phi}) = o_p(n_j^{-1}).
\]
By the Cauchy-Schwarz inequality, Assumptions 1(i), 1(iv), 3(iii), 43, 66 and 117,

\[
\left\| (nn_j)^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} W_{n,m}(Z_i) P_{k_j}(Z_i) Q_{n_j,k_j}^{-1} \sum_{j \in I_j} P_{k_j}(Z_i)(\phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i)^{\prime} \beta_{j,m,k_j}) \right\|^2 \\
\leq \left\| n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} W_{n,m}(Z_i) P_{k_j}(Z_i) \right\|^2 \left\| n_j^{-1} Q_{n_j,k_j}^{-1} \sum_{j \in I_j} P_{k_j}(Z_i)(\phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i)^{\prime} \beta_{j,m,k_j}) \right\|^2 \\
\leq \lambda_{\max}(Q_{n,k_j})(\sup_{z \in Z} (\lambda_{\max}(W_n(z))))^2 \sum_{i \in I} \left\| \gamma_{(1)} (Z_i, \theta_0) \right\|^2 n_j^{-1} \sum_{j \in I_j} \left\| P_{k_j}(Z_i)^{\prime} \beta_{j,m,k_j} - \phi_{j,m}(Z_i, \theta_0) \right\|^2 \\
= O_p(k_j^{-2r_\phi}) = o_p(n_j^{-1}). \tag{122}
\]

From the results in (119), (120), (121) and (122),

\[
n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} \hat{W}_n(Z_i) \hat{\phi}(Z_i, \theta_0) \\
= \sum_{m=1}^{d_y} n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} W_{n,m}(Z_i) P_{k_1}(Z_i)^{\prime} (P'_{n_1,k_1} P_{n_1,k_1})^{-1} \sum_{j \in I_j} P_{k_j}(Z_i) u_{1,m,i} \\
+ \sum_{m=1}^{d_y} n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} W_{n,m}(Z_i) P_{k_2}(Z_i)^{\prime} (P'_{n_2,k_2} P_{n_2,k_2})^{-1} \sum_{j \in I_j} P_{k_j}(Z_i) u_{2,m,i} \\
+ o_p(n_1^{-1} + n_2^{-1})
\]

which together with Lemma 11 and the definition of \( \phi_{w\theta,n} \) proves the lemma.

**Proof of Theorem 3.** By definition, \( \hat{\theta}_n \) satisfies the following first order condition

\[
n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \hat{\theta}_n)^{\prime} \hat{W}_n(Z_i) \hat{\phi}(Z_i, \hat{\theta}_n) = 0_{d_y}. \tag{123}
\]

Applying the first-order expansion row by row to (123), we get

\[
0_{d_y} = n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \theta_0)^{\prime} \hat{W}_n(Z_i) \hat{\phi}(Z_i, \theta_0) \\
+ n^{-1} \sum_{i \in I} \gamma_{(1)} (Z_i, \tilde{\theta}_n)^{\prime} \hat{W}_n(Z_i)(\hat{\phi}(Z_i, \tilde{\theta}_n)(\tilde{\theta}_n - \theta_0)) \\
+ n^{-1} \sum_{i \in I} \left( \gamma_{(2)} (Z_i, \tilde{\theta}_n)(\tilde{\theta}_n - \theta_0) \right)_{m=1, \ldots, d_y} \hat{W}_n(Z_i) \left( \hat{\phi}(Z_i, \tilde{\theta}_n) \right) , \tag{124}
\]

where \( \gamma_{(2)} (z, \theta) = \gamma_{1,m} (z, \theta) - \gamma_{2,m} (z, \theta) \), \( \tilde{\theta}_n \) is between \( \hat{\theta}_n \) and \( \theta_0 \) and it may be different across rows. For \( l = 1, \ldots, d_y \), let \( \tilde{\theta}_{n,l} \) denote the mean value in the \( l \)-th equation of (124). Since \( \tilde{\theta}_n \) is
By Lemma 4 and Lemma 8, consistent and \(d_g\) is finite, we have
\[
\max_{l=1,\ldots,d_g} \left\| \tilde{\theta}_{n,l} - \theta_0 \right\| = o_p(1). \tag{125}
\]

By (125), Lemma 9 and Assumption 3(iii),
\[
n^{-1} \sum_{i \in I} \left( \phi^{(1)}(Z_i, \tilde{\theta}_n) \right)^j \tilde{W}_n(Z_i) \left( \phi^{(1)}(Z_i, \tilde{\theta}_n) \right) = E \left[ \phi^{(1)}(Z, \theta_0)'W_n(Z)\phi^{(1)}(Z, \theta_0) \right] + o_p(1). \tag{126}
\]

By (126), Lemma 10 and Lemma 12, we can deduce from (124) that
\[
(-H_n + o_p(1))(\tilde{\theta}_n - \theta_0) = (nn_1)^{-1} \phi_{\theta,n}(I_m \otimes P_{n,k_1}Q_{1,n_1,k_1}^{-1}) \sum_{i \in I_l} u_{1,i} \otimes P_{k_1}(Z_i) \n
- (nn_2)^{-1} \phi_{\theta,n}(I_m \otimes P_{n,k_2}Q_{2,n_2,k_2}^{-1}) \sum_{i \in I_l} u_{2,i} \otimes P_{k_2}(Z_i) + o_p(n_1^{-1/2} + n_2^{-1/2}) \tag{127}
\]

which together with Assumption 2(ii) and Lemma 5 implies that
\[
\hat{\theta}_n - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2}). \tag{128}
\]

This shows the first claim of the theorem. By Assumption 2(ii), (127) and (128),
\[
\hat{\theta}_n - \theta_0 = -H_n^{-1} \frac{\phi_{\theta,n}}{n} \left[ \frac{\phi_{\theta,n}(I_m \otimes P_{n,k_1}Q_{1,n_1,k_1}^{-1})}{n_1} \sum_{i \in I_l} u_{1,i} \otimes P_{k_1}(Z_i) \right. \n
- \left. \frac{\phi_{\theta,n}(I_m \otimes P_{n,k_2}Q_{2,n_2,k_2}^{-1})}{n_2} \sum_{i \in I_l} u_{2,i} \otimes P_{k_2}(Z_i) \right] + o_p(n_1^{-1/2} + n_2^{-1/2}) \tag{129}
\]

By Lemma 4 and Lemma 8,
\[
n_j \Omega_{j,n} = \frac{\phi_{\theta,n}(I_m \otimes P_{n,k_j}Q_{n_j,k_j}^{-1})Q_{n_j,u_j}(I_m \otimes Q_{n_j,k_j}^{-1})P_{n,k_j}^{p'} \phi_{\theta,n}'}{n^2} \n
\geq \lambda_{\min}(Q_{n_j,u_j}) \frac{\phi_{\theta,n}(I_m \otimes P_{n,k_j}Q_{n_j,k_j}^{-2}P_{n,k_j}^{p'} \phi_{\theta,n}')}{n^2} \n
= \lambda_{\min}(Q_{n_j,u_j}) \sum_{m=1}^{d_g} \frac{\phi_{\theta,m,n}(P_{n,k_j}Q_{n_j,k_j}^{-2}P_{n,k_j}^{p'} \phi_{\theta,m,n})}{n^2} \n
\geq \lambda_{\min}(Q_{n_j,u_j}) \lambda_{\min}(Q_{n_j,k_j}) \sum_{m=1}^{d_g} n^{-1} \phi_{\theta,m,n}(P_{n,k_j}^{p'} P_{n,k_j} P_{n,k_j}^{p'} \phi_{\theta,m,n}) \n
= \frac{\lambda_{\min}(Q_{n_j,u_j}) \lambda_{\min}(Q_{n_j,k_j})}{(\lambda_{\max}(Q_{n_j,k_j}))^2} \left( E \left[ \phi^{(1)}(Z, \theta_0)'W_n(Z)\phi^{(1)}(Z, \theta_0) \right] + o_p(1) \right) \tag{130}
\]
which together with Assumptions 1(v), 2(ii) and 43 implies that

$$\lambda_{\min}(n_j \Omega_{j,n}) \geq C^{-1}$$  \hspace{1cm} (131)$$

with probability approaching 1. Without loss of generality, we set $I_1 = \{1, \ldots, n_1\}$ and $I_2 = \{n_1 + 1, \ldots, n\}$ where $n = n_1 + n_2$. Define

$$\omega_{i,n} = \left\{ \begin{array}{ll} -(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{-1/2} \phi_{w\theta,n}(I_m \otimes P_{n,k_1} Q_{n1,k_1}^{-1})(u_{1,i} \otimes P_{k_1}(Z_i)) & \text{for } 1 \leq i \leq n_1 \\
(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{-1/2} \phi_{w\theta,n}(I_m \otimes P_{n,k_2} Q_{n2,k_2}^{-1})(u_{2,i} \otimes P_{k_2}(Z_i)) & \text{for } n_1 < i \leq n \end{array} \right.$$  \hspace{1cm} (132)

Then by (129) and (131), we can write

$$(H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{-1/2}(\hat{\theta}_n - \theta_0) = \sum_{i=1}^{n} \omega_{i,n} + o_p(1).$$  \hspace{1cm} (133)

Let $\mathcal{F}_{0,n}$ be the $\sigma$-field generated \{Z_i\}_{i \in I}$, and $\mathcal{F}_{i,n}$ be the $\sigma$-field generated by \{(Z_i)_{i \in I}, \omega_{1,n}, \ldots, \omega_{i,n}\}$ for $i = 2, \ldots, n$. Then under Assumption 1(i), $E[\gamma_n' \omega_{i,n} | \mathcal{F}_{i-1,n}] = 0$ which means that \{\gamma_n' \omega_{i,n}\}_{i=1}^{n}$ is a martingale difference array. We next use the Martingale Central Limit Theorem (MCLT) to show the second claim of the theorem. There are two sufficient conditions to verify:

$$\sum_{i=1}^{n} E[(\gamma_n' \omega_{i,n})^2 | \mathcal{F}_{i-1,n}] \to_p 1; \text{ and}$$  \hspace{1cm} (134)

$$\sum_{i=1}^{n} E[(\gamma_n' \omega_{i,n})^2 I\{ |\gamma_n' \omega_{i,n}| > \varepsilon \} | \mathcal{F}_{i-1,n}] \to_p 0 \ \forall \varepsilon > 0.$$  \hspace{1cm} (135)

For ease of notations, we define $D_n = (H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n)^{-1/2}H_n^{-1}$. By definition, we have

$$\sum_{i=1}^{n} E[(\gamma_n' \omega_{i,n})^2 | \mathcal{F}_{i-1,n}] = \sum_{i=1}^{n} \gamma_n' E[\omega_{i,n} \omega_{i,n}' | \mathcal{F}_{i-1,n}] \gamma_n$$

$$= \gamma_n' D_n \phi_{w\theta,n}(I_m \otimes P_{n,k_1} Q_{n1,k_1}^{-1})Q_{n1,u_1}(I_m \otimes Q_{n1,k_1}^{-1} P_{n,k_1}' \phi_{w\theta,n}' D_n' \gamma_n$$

$$+ \gamma_n' D_n \phi_{w\theta,n}(I_m \otimes P_{n,k_2} Q_{n2,k_2}^{-1})Q_{n2,u_2}(I_m \otimes Q_{n2,k_2}^{-1} P_{n,k_2}' \phi_{w\theta,n}' D_n' \gamma_n$$

$$= \gamma_n' D_n(\Omega_{1,n} + \Omega_{2,n})D_n' \gamma_n = \gamma_n' \gamma_n = 1$$  \hspace{1cm} (136)
which proves (134). By the monotonicity of expectation,

\[
\sum_{i=1}^{n} E \left[ (\gamma'_{n}\omega_{i,n})^2 I \left\{ |\gamma'_{n}\omega_{i,n}| > \delta \right\} | F_{i-1,n} \right]
\]

\[
\leq \frac{1}{\varepsilon^2} \sum_{i=1}^{n} E \left[ (\gamma'_{n}\omega_{i,n})^4 | F_{i-1,n} \right]
\]

\[
= \frac{1}{\varepsilon^2} \sum_{i=1}^{n} E \left[ \gamma'_{n}D_n\phi_{w\theta,n}(I_m \otimes P_{n,k_1}Q^{-1}_{n_1,k_1})\left(u_{1,i} \otimes P_{k_1}(Z_i)\right) \right]^{4} \left| F_{i-1,n} \right|
\]

\[
+ \frac{1}{\varepsilon^2} \sum_{i=n_1+1}^{n} E \left[ \gamma'_{n}D_n\phi_{w\theta,n}(I_m \otimes P_{n,k_2}Q^{-1}_{n_2,k_2})\left(u_{2,i} \otimes P_{k_2}(Z_i)\right) \right]^{4} \left| F_{i-1,n} \right|
\]

(137)

For \( j = 1, 2 \) and \( m = 1, \ldots, d_g \), let \( u_{j,m,i} \) denote the \( m \)-th component of \( u_{j,i} \). By definition,

\[
\phi_{w\theta,n}(I_m \otimes P_{n,k_j}Q^{-1}_{n_j,k_j})(u_{j,i} \otimes P_{k_j}(Z_i)) = \sum_{m=1}^{d_g} \phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i)u_{j,m,i}.
\]

(138)

By Cauchy-Schwarz inequality,

\[
\left| \gamma'_{n}D_n\phi_{w\theta,n}(I_m \otimes P_{n,k_j}Q^{-1}_{n_j,k_j})(u_{j,i} \otimes P_{k_j}(Z_i)) \right|^4 \leq d_g^2 \sum_{m=1}^{d_g} \left| \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i)u_{j,m,i} \right|^4.
\]

(139)

By Assumption 2\( v \)) and the triangle inequality,

\[
(nn_j)^{-4} \sum_{i \in I_j} E \left[ \left| \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i)u_{j,m,i} \right|^4 \right] \left| F_{i-1,n} \right|
\]

\[
\leq C(nn_j)^{-4} \sum_{i \in I_j} \left| \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i) \right|^4
\]

\[
\leq Cn^{-4}n_j^{-3} \xi_{k_j} \times (\gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i)) \left( \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i) \right)^2
\]

\[
\times \left( \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}Q^{-1}_{n_j,k_j}P_{k_j}(Z_i) \right)^2
\]

\[
\leq \frac{C \xi_{k_j}^2}{\lambda_{\min}(Q_{n_j,k_j})^3n_j^2} \left( \gamma'_{n}D_n\phi_{w\theta,m,n}P_{n,k_j}P'_{n,k_j}(\phi'_{w\theta,m,n}D'_{n}\gamma_{n}) \right)^2
\]

(140)
By (61),
\[
\text{tr} \left( n^{-1} \phi_{w \theta, m, n} \phi'_{w \theta, m, n} \right) \leq \text{tr} \left( n^{-1} \phi_{w \theta, n} \phi'_{w \theta, n} \right) = O_p(1). \tag{141}
\]

By (131),
\[
\lambda_{\text{min}}(n_j(\Omega_{1, n} + \Omega_{2, n})) \geq \lambda_{\text{min}}(n_j \Omega_{j, n}) > C^{-1}. \tag{142}
\]

By Assumption 1(v) and (58),
\[
\lambda_{\text{max}}(H_n) \leq 2C. \tag{143}
\]

For any \( \gamma \in \mathbb{R}^{d_\theta} \), we have
\[
\frac{\gamma' D_n D' n_j \gamma}{n_j} = \frac{\gamma'(H_n(\Omega_{1, n} + \Omega_{2, n})^{-1}H_n)^{1/2}H_n^{-2}(H_n(\Omega_{1, n} + \Omega_{2, n})^{-1}H_n)^{1/2} \gamma}{n_j} \leq \frac{1}{(\lambda_{\text{min}}(H_n))^2} \frac{\gamma'(H_n(\Omega_{1, n} + \Omega_{2, n})^{-1}H_n)^{1/2} \gamma}{n_j} \leq \frac{\gamma' H_n^{2 \gamma}}{(\lambda_{\text{min}}(H_n))^2 \lambda_{\text{min}}(n_j(\Omega_{1, n} + \Omega_{2, n}))} \leq \frac{\gamma' H_n^{2 \gamma}}{(\lambda_{\text{max}}(H_n))^2 \lambda_{\text{min}}(n_j(\Omega_{1, n} + \Omega_{2, n}))},
\]
which combined with Assumption 2(ii), (142) and (143) implies that
\[
\lambda_{\text{max}}(n_j^{-1} D_n D'_n) \leq C. \tag{145}
\]

By (141) and (145),
\[
\frac{(\gamma' D_n D' n_j \phi_{w \theta, m, n} \phi'_{w \theta, m, n} D'_n \gamma n_j)^2}{n^2 n_j^2} \leq (\lambda_{\text{max}}(n^{-1} \phi_{w \theta, m, n} \phi'_{w \theta, m, n}))^2 (\lambda_{\text{max}}(n_j^{-1} D_n D'_n))^2 = O_p(1). \tag{146}
\]

Collecting the results in (140), (146) and (43), we deduce that
\[
(n n_j)^{-4} \sum_{i \in I_j} E \left[ \left| \gamma' D_n \phi_{w \theta, m, n} P_{n, k_j} Q_{n_j, k_j} P_{k_j}(Z_i) u_{j, m, i} \right|^4 \right| \mathcal{F}_{i-1, n} ] = O_p(\xi_j^2 n_j^{-1}), \tag{147}
\]
which together with Assumption 3(iii) and (139) implies that
\[
\frac{1}{\varepsilon^2} \sum_{i \in I_j} E \left[ \left| \gamma' D_n \phi_{w \theta, n} (I_m \otimes P_{n, k_j} Q_{n_j, k_j}) (u_{j, i} \otimes P_{k_j}(Z_i)) \right|^4 \right| \mathcal{F}_{i-1, n} ] = o_p(1) \tag{148}
\]
for any \( \varepsilon > 0. \) (135) now follows from (137) and (148). As a result, the asymptotic normality of \( \hat{\theta}_n \) follows by the MCLT.
B Proof of the Main Results in Section 4

Lemma 13. Under Assumptions (i), (ii), (iv), (v) and (iii), we have for $j = 1, 2$ and $m_1, m_2 = 1, \ldots, d_g$,

$$n^{-2} \phi_{w \theta, m_1, n} P_{n, k_j} Q^{-1}_{n, j, k_j} \left( n^{-1} \sum_{i \in I_j} \Sigma_{j, m_1, m_2}(Z_i) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right) Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_2, n}$$

$$= E \left[ \Sigma_{j, m_1, m_2}(Z) \phi^{(1)}(Z, \theta_0) W_{n, m_1}(Z) W_{n, m_2}(Z) \Delta \phi^{(1)}(Z, \theta_0) \right] + o_p(1),$$

where $\Sigma_{j, m_1, m_2}(Z) = E[u_{j, m_1, m_2} | Z]$.

Proof of Lemma 13. For $m = 1, \ldots, d_g$ and $l = 1, \ldots, d_\theta$, let $\phi_{w \theta, m, n}$ denote the $l$-th row of $\phi_{w \theta, m, n}$. Define

$$\tilde{\phi}^{(1)}_m(z) = n^{-1} P_{k_j}(z)^{Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_1, n}}$$

and

$$\tilde{\phi}^{(1)}_{m,l}(z) = n^{-1} P_{k_j}(z)^{Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_1, n}}$$

(149)

where $\tilde{\phi}^{(1)}_{m,l}(z)$ is the $l$-th component in $\tilde{\phi}^{(1)}_m(z)$. We can write

$$n^{-2} \phi_{w \theta, m_1, n} P_{n, k_j} Q^{-1}_{n, j, k_j} \left( n^{-1} \sum_{i \in I_j} \Sigma_{j, m_1, m_2}(Z_i) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right) Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_2, n}$$

$$= n^{-1} \sum_{i \in I_j} \Sigma_{j, m_1, m_2}(Z_i) \tilde{\phi}^{(1)}_{m_1}(Z_i) \tilde{\phi}^{(1)}_{m_2}(Z_i).$$

(150)

Consider any $m = 1, \ldots, d_g$ and any $l = 1, \ldots, d_\theta$. Define

$$\tilde{\phi}^{(1)}_m(z) = n^{-1} P_{k_j}(z)^{Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_1, n}}$$

and

$$\tilde{\phi}^{(1)}_{m,l}(z) = n^{-1} P_{k_j}(z)^{Q^{-1}_{n, j, k_j} P'_{n, k_j} \phi_{w \theta, m_1, n}}$$

(151)

By (149) and (150),

$$n^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m,l}(Z_i) \right|^2 = n_j n^{-2} \phi_{w \theta, m_1, n} P_{n, k_j} (P'_{n, k_j} P_{n, k_j})^{-1} P'_{n, k_j} \phi_{w \theta, m_1, n}$$

$$\leq \frac{\lambda_{\max}(Q_{n, k_j})}{\lambda_{\min}(Q_{n, k_j})} n^{-1} \phi_{w \theta, m_1, n} P_{n, k_j} (P'_{n, k_j} P_{n, k_j})^{-1} P'_{n, k_j} \phi_{w \theta, m_1, n}$$

$$\leq \frac{\lambda_{\max}(Q_{n, k_j})}{\lambda_{\min}(Q_{n, k_j})} \frac{\phi_{w \theta, m_1, n} \phi_{w \theta, m_1, n} P_{n, k_j} (P'_{n, k_j} P_{n, k_j})^{-1} P'_{n, k_j} \phi_{w \theta, m_1, n}}{n} = o_p(1).$$

Similarly, we can show that

$$n^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m,l}(Z_i) \right|^2 = o_p(1).$$

(152)
By Assumption $1(iv)$, $58$ and the Markov inequality,

$$n_j^{-1} \sum_{i \in I_j} \left| \phi^{(1)}_{\theta_i}(Z_i, \theta_0)'W_{n,m}(Z_i) \right|^2 = n_j^{-1} \sum_{i \in I_j} \phi^{(1)}_{\theta_i}(Z_i, \theta_0)'W_{n,m}(Z_i)W_{n,m}(Z_i)'\phi^{(1)}_{\theta_i}(Z_i, \theta_0) \leq \sup_{z \in \mathcal{Z}}(\lambda_{\text{max}}(W_n(z)))^2 n_j^{-1} \sum_{i \in I_j} \phi^{(1)}_{\theta_i}(Z_i, \theta_0)'\phi^{(1)}_{\theta_i}(Z_i, \theta_0) = O_p(1). \quad (153)$$

By Assumptions $2(iv)$ and $3(iii)$, $42$, $43$ and $57$,

$$n_j^{-2} \sum_{i \in I_j} \left| P_{k_j}(Z_i)'(Q_{n,j}^{-1} - Q_{k_j}^{-1})P'_{n,k_j} \phi'_{\omega \theta_i,m,n} \right|^2 = n_j^{-2} \phi_{\omega \theta_i,m,n} P_{n,k_j} Q_{n,j}^{-1}(Q_{n,j,k_j} - Q_{k_j})'Q_{n,j,k_j}^{-1} P'_{n,k_j} \phi'_{\omega \theta_i,m,n} \leq \frac{\lambda_{\text{max}}((Q_{n,j,k_j} - Q_{k_j})^2)}{\lambda_{\text{min}}(Q_{n,j,k_j}) \lambda_{\text{min}}(Q_{k_j})^2} n_j^{-2} \phi_{\omega \theta_i,m,n} P_{n,k_j} P'_{n,k_j} \phi'_{\omega \theta_i,m,n} \leq \frac{\lambda_{\text{max}}((Q_{n,j,k_j} - Q_{k_j})^2) \lambda_{\text{max}}(Q_{n,k_j})}{\lambda_{\text{min}}(Q_{n,j,k_j}) \lambda_{\text{min}}(Q_{k_j})^2} n_j^{-1} \phi_{\omega \theta_i,m,n} P_{n,k_j} (P_{n,k_j} P'_{n,k_j})^{-1} P'_{n,k_j} \phi'_{\omega \theta_i,m,n} \leq \frac{\lambda_{\text{max}}((Q_{n,j,k_j} - Q_{k_j})^2) \lambda_{\text{max}}(Q_{n,k_j})}{\lambda_{\text{min}}(Q_{n,j,k_j}) \lambda_{\text{min}}(Q_{k_j})^2} n_j^{-1} \phi_{\omega \theta_i,m,n} \phi'_{\omega \theta_i,m,n} = O_p(\xi^2_{k_j} \log(k_j)n_j^{-1}) = o_p(1). \quad (154)$$

Similarly, we can show that

$$n_j^{-2} \sum_{i \in I_j} \left| P_{k_j}(Z_i)'(Q_{n,j}^{-1} - Q_{k_j}^{-1})P'_{n,k_j} \phi_{\omega \theta_i,m,n} \right|^2 = o_p(1). \quad (155)$$

By $154$ and $155$,

$$n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}_{m,I}^{(1)}(Z_i) - \phi_{m,I}^{(1)}(Z_i) \right|^2 = o_p(1). \quad (156)$$
By Assumption 2(v), (151), (152), (156), the triangle inequality and the Cauchy-Schwarz inequality,

\[
\left| n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2}(Z_i) \left( \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) \tilde{\phi}^{(1)}_{m_2,l_2}(Z_i) - \hat{\phi}^{(1)}_{m_1,l_1}(Z_i) \hat{\phi}^{(1)}_{m_2,l_2}(Z_i) \right) \right| \\
\leq n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2}(Z_i) \left( \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) \left( \tilde{\phi}^{(1)}_{m_2,l_2}(Z_i) - \hat{\phi}^{(1)}_{m_2,l_2}(Z_i) \right) \right) \\
+ n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2}(Z_i) \left( \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) - \hat{\phi}^{(1)}_{m_1,l_1}(Z_i) \right) \tilde{\phi}^{(1)}_{m_2,l_2}(Z_i) \\
\leq C \left( n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) \right|^2 \right)^{1/2} \left( n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m_2,l_2}(Z_i) - \hat{\phi}^{(1)}_{m_2,l_2}(Z_i) \right|^2 \right)^{1/2} \\
+ C \left( n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m_2,l_2}(Z_i) \right|^2 \right)^{1/2} \left( n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) - \hat{\phi}^{(1)}_{m_1,l_1}(Z_i) \right|^2 \right)^{1/2} = o_p(1), \quad (157)
\]

which implies that

\[
n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2}(Z_i) \left( \tilde{\phi}^{(1)}_{m_1,l_1}(Z_i) \hat{\phi}^{(1)}_{m_2,l_2}(Z_i) - \hat{\phi}^{(1)}_{m_1,l_1}(Z_i) \hat{\phi}^{(1)}_{m_2,l_2}(Z_i) \right) = o_p(1). \quad (158)
\]

By (77) and the Markov inequality,

\[
n_j^{-1} \sum_{i \in I_j} \left| \tilde{\phi}^{(1)}_{m_1}(Z_i) - W_{n,m}(Z_i)' \phi^{(1)}_{m}(Z_i) \right|^2 \\
\leq n_j^{-1} \sum_{i \in I_j} \left| W_{n,m}(Z_i)' \phi^{(1)}_{m}(Z_i) - P_{kj}(Z_i)'(P'_{n,kj} P_{n,kj})^{-1} P'_{n,kj} \phi^{(1)}_{w_{\theta,n}} \right|^2 = o_p(1). \quad (159)
\]
By the triangle inequality and the Cauchy-Schwarz inequality,

\[
\begin{align*}
& n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2} (Z_i) \left( \phi_{m_1,1}^{(1)} (Z_i) \phi_{m_2,2}^{(1)} (Z_i) - \phi_{\theta_1}^{(1)} (Z_i) W_{n,m_1} (Z_i) \phi_{\theta_2}^{(1)} (Z_i) W_{n,m_2} (Z_i) \right) \\
\leq \ & n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2} (Z_i) \phi_{m_1,1}^{(1)} (Z_i) \phi_{m_2,2}^{(1)} (Z_i) \\
& + n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2} (Z_i) \phi_{\theta_2}^{(1)} (Z_i) W_{n,m_2} (Z_i) \phi_{\theta_1}^{(1)} (Z_i) W_{n,m_1} (Z_i) \\
\leq \ & C \left( n_j^{-1} \sum_{i \in I_j} \phi_{m_1,1}^{(1)} (Z_i) \phi_{m_2,2}^{(1)} (Z_i) \right)^2 \left( n_j^{-1} \sum_{i \in I_j} \phi_{\theta_2}^{(1)} (Z_i) \phi_{\theta_1}^{(1)} (Z_i) \right)^2 \right)^{1/2} \\n& + C \left( n_j^{-1} \sum_{i \in I_j} \phi_{\theta_2}^{(1)} (Z_i) W_{n,m_2} (Z_i) \right)^2 \left( n_j^{-1} \sum_{i \in I_j} \phi_{\theta_1}^{(1)} (Z_i) W_{n,m_1} (Z_i) \right)^2 \right)^{1/2} = o_p(1),
\end{align*}
\]

(160)

where the equality is by (152), (153) and (159). By Assumption 2(iii) and Jensen’s inequality,

\[
E \left[ \| \partial \phi_{j,m} (Z, \theta_0) / \partial \theta_1 \|^4 \right] = E \left[ \| E \left[ \partial g_{j,m} (Y_j, \theta_0) / \partial \theta_1 | Z \right] \|^4 \right] \leq d_0 E \left[ \| \partial g_{j,m} (Y_j, \theta_0) / \partial \theta_1 \|^4 \right] \leq C.
\]

(161)

By Assumptions 1(i), 1(v), 2(v), (161) and the Markov inequality,

\[
\begin{align*}
& n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2} (Z) \phi_{j,m}^{(1)} (Z, \theta_0) W_{n,m_1} (Z) W_{n,m_2} (Z) \phi_{j,m}^{(1)} (Z, \theta_0) \\
= \ & E \left[ \Sigma_{j,m_1,m_2} (Z) \phi_{j,m}^{(1)} (Z, \theta_0) W_{n,m_1} (Z) W_{n,m_2} (Z) \phi_{j,m}^{(1)} (Z, \theta_0) \right] + o_p(1)
\end{align*}
\]

(162)

for any \(m_1, m_2 = 1, \ldots, d_g\). The claim of the Lemma follows by (150), (158), (160) and (162). 

\textbf{Proof of Lemma 7} By Lemma 13

\[
\begin{align*}
n_j \Omega_{j,n} = \ & n^{-2} \phi_{w \theta,n} (I_m \otimes P_{n,k_j} Q_{n_j,k_j}^{-1}) Q_{n_j,u_j} (I_m \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}') \phi_{w \theta,n} \\
= \ & \sum_{m_1,m_2 = 1}^{d_g} n^{-2} \phi_{w \theta,m_1,m_2} P_{n,k_j} Q_{n_j,k_j}^{-1} \left( n_j^{-1} \sum_{i \in I_j} \Sigma_{j,m_1,m_2} (Z_i) P_{k_j} (Z_i) P_{k_j}' (Z_i) \right) Q_{n_j,k_j}^{-1} P_{n,k_j} \phi_{w \theta,m_2,n} \\
= \ & \sum_{m_1,m_2 = 1}^{d_g} E \left[ \Sigma_{j,m_1,m_2} (Z) \phi_{m_1}^{(1)} (Z, \theta_0) W_{n,m_1} (Z) W_{n,m_2} (Z) \phi_{m_2}^{(1)} (Z, \theta_0) \right] + o_p(1) \\
= \ & E \left[ \phi_{j}^{(1)} (Z, \theta_0) W_{n} (Z) \Sigma_j (Z) W_{n} (Z) \phi_{j}^{(1)} (Z, \theta_0) \right] + o_p(1)
\end{align*}
\]

(163)
By Assumptions 1(iv), 2(ii) and 2(v),

$$\lambda_{\min} \left( E \left[ \phi^{(1)}(Z, \theta_0)'W_n(Z)\Sigma_j(Z)W_n(Z)\phi^{(1)}(Z, \theta_0) \right] \right) \geq C^{-1} \quad (164)$$

which together with (163) implies that

$$n_j \Omega_{j,n} = E \left[ \phi^{(1)}(Z, \theta_0)'W_n(Z)\Sigma_j(Z)W_n(Z)\phi^{(1)}(Z, \theta_0) \right] (1 + o_p(1)). \quad (165)$$

The claim of the lemma follows from (165).

**Proof of Theorem 3.** For any $W_n(\cdot)$, define

$$A_n(Z, W_n) = H_n^{-1}\phi^{(1)}(Z, \theta_0)'W_n(Z) - (H_n^*)^{-1}\phi^{(1)}(Z, \theta_0)'W_n^*(Z). \quad (166)$$

Then there is

$$V_{n,\theta} - V_{n,\theta}^* = E \left[ A_n(Z, W_n)(n_1^{-1}\Sigma_1(Z) + n_2^{-1}\Sigma_2(Z))A_n(Z, W_n)' \right] \geq 0, \quad (167)$$

for any $n_1$ and any $n_2$, where the inequality is by the fact that $n_1^{-1}\Sigma_1(Z) + n_2^{-1}\Sigma_2(Z)$ is strictly positive definite and $A_n(Z, W_n)(n_1^{-1}\Sigma_1(Z) + n_2^{-1}\Sigma_2(Z))A_n(Z, W_n)'$ is a positive semi-definite matrix for any $Z$.

**Proof of Lemma 2.** By definition

$$(W_n^*(z))^{-1} = (n_1^{-1} + n_2^{-1})^{-1}(n_1^{-1}\Sigma_1(z) + n_2^{-1}\Sigma_2(z)).$$

Then for any $n_1$ and any $n_2$, by Assumption 2(v),

$$\sup_{z \in \mathcal{Z}} \lambda_{\max}((W_n^*(z))^{-1}) \leq \sup_{z \in \mathcal{Z}} \lambda_{\max}(\Sigma_1(z)) + \sup_{z \in \mathcal{Z}} \lambda_{\max}(\Sigma_2(z)) \leq 2C \quad (168)$$

and

$$\inf_{z \in \mathcal{Z}} \lambda_{\min}((W_n^*(z))^{-1}) \geq 2^{-1} \min \left\{ \inf_{z \in \mathcal{Z}} \lambda_{\min}(\Sigma_1(z)), \inf_{z \in \mathcal{Z}} \lambda_{\min}(\Sigma_1(z)) \right\} \geq (2C)^{-1}. \quad (169)$$

Since $\lambda_{\max}(W_n^*(z)) = (\lambda_{\min}((W_n^*(z))^{-1}))^{-1}$ and $\lambda_{\min}(W_n^*(z)) = (\lambda_{\max}((W_n^*(z))^{-1}))^{-1}$, the claim of the lemma follows from (168) and (169).
C Proof of the Main Results in Section 5

Lemma 14. Under Assumptions 1, 2, 3, and 4(i), we have for $m = 1, \ldots, d_g$

\[ n_j^{-1} \sum_{i \in I_j} \left( \hat{\phi}_{j,m}(Z_i, \tilde{\theta}_n) - \phi_{j,m}(Z_i, \theta_0) \right)^2 = O_p(n_1^{-1} + n_2^{-1} + k_j n_j^{-1} + k_j^{-2r_0}) \]  \hspace{1cm} (170)

and moreover,

\[ \sup_{z \in Z} \left| \hat{\phi}_{j,m}(z, \tilde{\theta}_n) - \phi_{j,m}(z, \theta_0) \right| = O_p(\xi_k (n_1^{-1/2} + n_2^{-1/2} + k_j^{1/2} n_j^{-1/2} + k_j^{-r_0})) \]  \hspace{1cm} (171)

for $m = 1, \ldots, d_g$.

Proof of Lemma 14. By definition,

\[ \hat{\phi}_{j,m}(z, \tilde{\theta}_n) = P'_{k_j}(z)(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} g_{j,m,n_j}(\hat{\theta}_n) \]

where $g_{j,m,n_j}(\theta) = ((g_{j,m}(Z_i, \theta))_{i \in I_j})'$ for any $\theta \in \Theta$. By the first-order Taylor expansion,

\[ g_{j,m}(Y_{j,i}, \theta) - g_{j,m}(Y_{j,i}, \theta_0) = g_{j,m}^{(1)}(Y_{j,i}, \tilde{\theta}_n)(\theta - \theta_0) \]

where $\tilde{\theta}_n$ is between $\theta$ and $\theta_0$ and it may be different across rows. By the Cauchy-Schwarz inequality,

\[ |g_{j,m}(Y_{j,i}, \theta) - g_{j,m}(Y_{j,i}, \theta_0)| \leq d_\theta \sup_{\theta \in \Theta} \left\| g_{j,m}^{(1)}(Y_{j,i}, \theta) \right\| \|	heta - \theta_0\|, \]  \hspace{1cm} (172)

which together with Assumption 4(i) and 7 in Theorem 2 implies that

\[ n_j^{-1} \sum_{i \in I_j} \left| g_{j,m}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m}(Y_{j,i}, \theta_0) \right|^2 \leq d_\theta^2 \left\| \tilde{\theta}_n - \theta_0 \right\|^2 \sup_{\theta \in \Theta} n_j^{-1} \sum_{i \in I_j} \left\| g_{j,m}^{(1)}(Y_{j,i}, \theta) \right\|^2 = O_p(n_1^{-1} + n_2^{-1}). \]  \hspace{1cm} (173)

Since $P_{n_j,k_j}(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j}$ is an idempotent matrix

\[ n_j^{-1} \sum_{i \in I_j} \left| \hat{\phi}_{j,m}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m}(Z_i, \theta_0) \right|^2 \]

\[ = n_j^{-1} \left( g_{j,m,n_j}(\tilde{\theta}_n) - g_{j,m,n_j}(\theta_0) \right)' P_{n_j,k_j}(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \left( g_{j,m,n_j}(\tilde{\theta}_n) - g_{j,m,n_j}(\theta_0) \right) \]

\[ \leq n_j^{-1} \sum_{i \in I_j} \left| g_{j,m}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m}(Y_{j,i}, \theta_0) \right|^2 = O_p(n_1^{-1} + n_2^{-1}). \]  \hspace{1cm} (174)

The claim in (170) now follows from (69) and (174).
By the Cauchy-Schwarz inequality,

\[
\left| \hat{\phi}_{j,m}(z, \hat{\theta}_n) - \hat{\phi}_{j,m}(z, \theta_0) \right| \\
= \left| P'_{k_j}(z) (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \left( g_{j,m,n_j}(\hat{\theta}_n) - g_{j,m,n_j}(\theta_0) \right) \right| \\
\leq \left\| P_{k_j}(z) \right\| (\lambda_{\min}(Q_{n_j,k_j}))^{-1/2} \left| n_j^{-1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} P'_{n_j,k_j} \left( g_{j,m,n_j}(\hat{\theta}_n) - g_{j,m,n_j}(\theta_0) \right) \right|
\]

\[
\leq \xi_{k_j} (\lambda_{\min}(Q_{n_j,k_j}))^{-1/2} \left( n_j^{-1} \sum_{i \in I_j} \left| g_{j,m}(Z_i, \hat{\theta}_n) - g_{j,m}(Z_i, \theta_0) \right|^2 \right)^{1/2}
\]

(175)

where the second inequality is by the definition of \( \xi_{k_j} \) and the fact that \( P_{n_j,k_j}(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \) is an idempotent matrix. By (43), (173) and (175),

\[
\sup_{z \in Z} \left| \hat{\phi}_{j,m}(z, \hat{\theta}_n) - \hat{\phi}_{j,m}(z, \theta_0) \right| = O_p(\xi_{k_j}(n_1^{-1/2} + n_2^{-1/2})).
\]

(176)

By Assumptions 1(i) and 2(v),

\[
E \left[ \left( n_j^{-1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} \sum_{i \in I_j} P_{k_j}(Z_i) (g_{j,m}(Y_{j,i}, \theta_0) - \phi_{j,m}(Z_i, \theta_0)) \right)^2 \right] \leq n_j^{-1} \sum_{i \in I_j} \left( \phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i) \beta_{\phi_{j,m,k_j}} \right)^2 = O_p(k_j^{-2r_{\phi}}).
\]

(177)

which together with the Markov inequality implies that

\[
\left| n_j^{-1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} P'_{n_j,k_j} \left( g_{j,m,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0) \right) \right| = O_p(k_j^{1/2} n_j^{-1/2}),
\]

(178)

where \( \phi_{j,m,n_j}(\theta_0) = (\phi_{j,m}(Z_i, \theta_0))_{i \in I_j} \)' . Let \( \phi_{j,m,k_j,n_j}(\theta_0) = ((P_{k_j}(Z_i) \beta_{\phi_{j,m,k_j}})_{i \in I_j} \)' . Then by Assumption 3(i),

\[
\left( n_j^{-1/2} (P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} P'_{n_j,k_j} \left( \phi_{j,m,k_j,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0) \right) \right)^2 \leq n_j^{-1} \sum_{i \in I_j} (\phi_{j,m}(Z_i, \theta_0) - P_{k_j}(Z_i) \beta_{\phi_{j,m,k_j}})^2 = O_p(k_j^{-2r_{\phi}}).
\]

(179)

By definition,

\[
\hat{\phi}_{j,m}(z, \theta_0) - \phi_{j,m}(z, \theta_0) = P_{k_j}(z)' (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \left( g_{j,m,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0) \right) \]

\[
+ P_{k_j}(z)' (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \left( \phi_{j,m,n_j}(\theta_0) - \phi_{j,m,k_j,n_j}(\theta_0) \right) \]

\[
+ P_{k_j}(z)' \beta_{\phi_{j,m,k}} - \phi_{j,m}(Z_i, \theta_0).
\]

(180)
By the Cauchy-Schwarz inequality, Assumption 3(i), (43), (178), (179) and (180)

\[
\sup_{z \in \mathbb{Z}} \left| \hat{\phi}_{j,m}(z, \theta_0) - \phi_{j,m}(z, \theta_0) \right| \\
\leq \frac{\sup_{z \in \mathbb{Z}} \left\| P_{k_j}(z) \right\|}{(\lambda_{\min}(Q_{n,j,k_j}))^{1/2}} \left\| n_{j}^{-1/2}(P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} P'_{n_j,k_j} (g_{j,m,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0)) \right\| \\
+ \frac{\sup_{z \in \mathbb{Z}} \left\| P_{k_j}(z) \right\|}{(\lambda_{\min}(Q_{n,j,k_j}))^{1/2}} \left\| n_{j}^{-1/2}(P'_{n_j,k_j} P_{n_j,k_j})^{-1/2} P'_{n_j,k_j} (\phi_{j,m,k_j,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0)) \right\| \\
+ \sup_{z \in \mathbb{Z}} \left| P_{k_j}(z)' \beta_{\phi,m,k} - \phi_{j,m}(z, \theta_0) \right| = O_p(\xi_k (k_j^{-1/2} n_j^{-1/2} + k_j^{-\tau_0}))
\]  

which together with (176) and the triangle inequality proves the second claim of the lemma. ■

**Lemma 15.** Under Assumptions 1, 2, 3 and 4, we have

(i) \( \hat{H}_n = H_n + o_p(1) \);

(ii) \( n^{-1}(\hat{\phi}_{w_0,m,n} - \phi_{w_0,m,n}) P_{n,k_j} = O_p(n_1^{-1/2} + n_2^{-1/2} + \delta_{w,n}) \) for \( m = 1, \ldots, d_g \);

(iii) \( \| \hat{Q}_{n,j,u_j} - Q_{n,j,u_j} \| = O_p(\xi_k k_j^{-1/2} (n_1^{-1/2} + n_2^{-1/2})) \);

(iv) \( n_j(\hat{\Omega}_{j,n} - \Omega_{j,n}) = O_p(\xi_k k_j^{-1/2} (n_1^{-1/2} + n_2^{-1/2}) + \delta_{w,n}) \);

(v) \( \hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} \hat{H}_n(H_n^{-1}(\Omega_{1,n} + \Omega_{2,n})H_n^{-1}) - I_{d_0} = o_p(1) \).

**Proof of Lemma 15** (i) Since \( \| \hat{\theta}_n - \theta_0 \| = O_p(n_1^{-1/2} + n_2^{-1/2}) \), by Lemma 9,

\[
\hat{H}_n = H_n + o_p(1),
\]

which proves the first claim of the lemma.

(ii) By definition,

\[
\begin{align*}
n^{-1}(\hat{\phi}_{w_0,m,n} - \phi_{w_0,m,n}) P_{n,k_j} \\
= \sum_{i \in I} \left( \hat{\phi}^{(1)}(Z_i, \hat{\theta}_n) - \hat{\phi}^{(1)}(Z_i, \theta_0) \right)' \hat{W}_{n,m}(Z_i) P_{k_j}(Z_i)' \\
- n^{-1} \sum_{i \in I} \hat{\phi}^{(1)}(Z_i, \theta_0)'(\hat{W}_{n,m}(Z_i) - W_{n,m}(Z_i)) P_{k_j}(Z_i)' .
\end{align*}
\]  

(182)
By the Cauchy-Schwarz inequality,
\[
\left\| n^{-1} \sum_{i \in I} \left( \phi^{(1)}(Z_i, \hat{\theta}_n) - \phi^{(1)}(Z_i, \theta_0) \right)' \hat{W}_{n,m}(Z_i)P_{kj}(Z_i) \right\|^2 \\
\leq \lambda_{\max}(Q_{n,k}) n^{-1} \sum_{i \in I} \left\| \left( \phi^{(1)}(Z_i, \hat{\theta}_n) - \phi^{(1)}(Z_i, \theta_0) \right)' \hat{W}_{n,m}(Z_i) \right\|^2 \\
\leq \lambda_{\max}(Q_{n,k}) \sup_{z \in Z} \lambda_{\max}(\hat{W}_{n}(z)) n^{-1} \sum_{i \in I} \left\| \phi^{(1)}(Z_i, \hat{\theta}_n) - \phi^{(1)}(Z_i, \theta_0) \right\|^2 \\
= O_p(n_1^{-1} + n_2^{-1})
\]

where the equality is by Assumption [3 iii), (7) in Theorem 2 (33), (43) and (88). Similarly we can show that

\[
\left\| n^{-1} \sum_{i \in I} \phi^{(1)}(Z_i, \theta_0)'(\hat{W}_{n,m}(Z_i) - W_{n,m}(Z_i))P_{kj}(Z_i) \right\|^2 \\
\leq \lambda_{\max}(Q_{n,k}) n^{-1} \sum_{i \in I} \left\| \phi^{(1)}(Z_i, \theta_0)'(\hat{W}_{n,m}(Z_i) - W_{n,m}(Z_i)) \right\|^2 \\
\leq \lambda_{\max}(Q_{n,k}) \sup_{z \in Z} \left\| \hat{W}_{n}(z) - W_{n}(z) \right\|^2 n^{-1} \sum_{i \in I} \left\| \phi^{(1)}(Z_i, \theta_0) \right\|^2 = O_p(\delta_{w,n}^2),
\]

where the equality is by Assumptions [4 iv), Lemma 7 (43) and (90). The claim of the lemma follows from (182), (183) and (184).

(iii) To prove this claim, it is sufficient to show that

\[
n_j^{-1} \sum_{i \in I_j} (\hat{u}_{j,m_1,i} - \hat{u}_{j,m_2,i} - \Sigma_{j,m_1,m_2}(Z_i))P_{kj}(Z_i)P'_{kj}(Z_i) = O_p(\xi_{k,j}^{1/2}(n_1^{-1/2} + n_2^{-1/2}))
\]

for any \(m_1, m_2 = 1, \ldots, d_g\), because \(d_g\) is finite and

\[
\left\| \hat{Q}_{n,u_j} - Q_{n,u_j} \right\|^2 = \sum_{m_1, m_2=1}^{d_g} n_j^{-1} \sum_{i \in I_j} (\hat{u}_{j,m_1,i} - \hat{u}_{j,m_2,i} - \Sigma_{j,m_1,m_2}(Z_i))P_{kj}(Z_i)P'_{kj}(Z_i) \right\|^2.
\]

Recall that for any \(m = 1, \ldots, d_g\)

\[
\hat{u}_{j,m,i} = u_{j,m,i} + (g_{j,m}(Y_{j,i}, \hat{\theta}_n) - g_{j,m}(Y_{j,i}, \theta_0)) - (\hat{\phi}_{j,m}(Z_i, \hat{\theta}_n) - \phi_{j,m}(Z_i, \theta_0)).
\]

57
By Assumptions 1(ii) and 2(v), and (43)

\[
E \left[ \left\| n_j^{-1} \sum_{i \in I_j} (u_{j,m_1,i}u_{j,m_2,i} - \Sigma_{j,m_1,m_2} (Z_i)) P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \right] \{ Z_i \}_{i \in I} \leq n_j^{-2} \sum_{i \in I_j} E \left[ u_{j,m_1,i}^2 \right] \left\| P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \leq C \xi_{kj}^2 \text{tr} (Q_{n_j,k_j} n_j^{-1}) = \mathcal{O}_p (\xi_{kj}^2 k_j n_j^{-1}).
\]

which together with the Markov inequality implies that

\[
n_j^{-1} \sum_{i \in I_j} (u_{j,m_1,i}u_{j,m_2,i} - \Sigma_{j,m_1,m_2} (Z_i)) P_{k_j} (Z_i) P'_{k_j} (Z_i) = \mathcal{O}_p (\xi_{kj}^2 k_j^{-1/2} n_j^{-1/2}).
\]

By Assumption 2(v) and (43),

\[
n_j^{-1} \sum_{i \in I_j} E \left[ \left\| u_{j,m_1,i} P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \right] \{ Z_i \}_{i \in I} = n_j^{-1} \sum_{i \in I_j} E \left[ u_{j,m_1,i}^2 \right] \left\| P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \leq C \xi_{kj}^2 n_j^{-1} \sum_{i \in I_j} \left\| P_{k_j} (Z_i) \right\|^2 = \mathcal{O}_p (\xi_{kj}^2 k_j),
\]

which together with the Markov inequality implies that

\[
n_j^{-1} \sum_{i \in I_j} \left\| u_{j,m_1,i} P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 = \mathcal{O}_p (\xi_{kj}^2 k_j).
\]

By the Cauchy-Schwarz inequality, (173) and (190),

\[
\left\| n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i} (g_{j,m_2} (Y_{j,i}, \hat{\theta}_n) - g_{j,m_2} (Y_{j,i}, \theta_0)) P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \leq n_j^{-1} \sum_{i \in I_j} \left\| u_{j,m_1,i} P_{k_j} (Z_i) P'_{k_j} (Z_i) \right\|^2 \times n_j^{-1} \sum_{i \in I_j} (g_{j,m_2} (Y_{j,i}, \hat{\theta}_n) - g_{j,m_2} (Y_{j,i}, \theta_0))^2 = \mathcal{O}_p (\xi_{kj}^2 k_j (n_1^{-1} + n_2^{-1})).
\]
By the Cauchy-Schwarz inequality, (174) and (190),

\[
\left\| n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i} (\hat{\phi}_{j,m_2}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0)) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right\|^2 
\leq n_j^{-1} \sum_{i \in I_j} \left\| u_{j,m_1,i} P_{k_j}(Z_i) P'_{k_j}(Z_i) \right\|^2 
\times n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_2}(Y_{j,i}, \widehat{\theta}_n) - \hat{\phi}_{j,m_2}(Y_{j,i}, \theta_0))^2 = O_p(\xi_k^2 k_j(n_1^{-1} + n_2^{-1})).
\]  

(192)

By the Cauchy-Schwarz inequality and (174),

\[
\left\| n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_1}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_1}(Z_i, \theta_0)) (\hat{\phi}_{j,m_2}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0)) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right\|^2 
\leq n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_1}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_1}(Z_i, \theta_0))^2 \| P_{k_j}(Z_i) \|^2 
\times n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_2}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0))^2 \| P_{k_j}(Z_i) \|^2 
\leq \xi_k^4 n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_1}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_1}(Z_i, \theta_0))^2 
\times n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_2}(Z_i, \widehat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0))^2 = O_p(\xi_k^4 (n_1^{-2} + n_2^{-2})).
\]  

(193)

By the Cauchy-Schwarz inequality and (173),

\[
\left\| n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0)) (g_{j,m_2}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right\|^2 
\leq n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0))^2 \| P_{k_j}(Z_i) \|^2 
\times n_j^{-1} \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^2 \| P_{k_j}(Z_i) \|^2 
\leq \xi_k^4 n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0))^2 
\times n_j^{-1} \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}, \widehat{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^2 = O_p(\xi_k^4 (n_1^{-2} + n_2^{-2})).
\]  

(194)
By the Cauchy-Schwarz inequality, (173) and (174),

\[
\left| \sum_{i \in I_j} \left( g_{j,m_1}(Y_{j,i}, \hat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0) \right) \left( \hat{\phi}_{j,m_2}(Z_i, \hat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0) \right) P_{k_j}(Z_i) P'_{k_j}(Z_i) \right|^2 \\
\leq \sum_{i \in I_j} \left( g_{j,m_1}(Y_{j,i}, \hat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0) \right)^2 \left\| P_{k_j}(Z_i) \right\|^2 \\
\times \sum_{i \in I_j} \left( \hat{\phi}_{j,m_2}(Z_i, \hat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0) \right)^2 \left\| P_{k_j}(Z_i) \right\|^2 \\
\leq \xi_k^4 n_j^{-1} \sum_{i \in I_j} \left( g_{j,m_1}(Y_{j,i}, \hat{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0) \right)^2 \\
\times \sum_{i \in I_j} \left( \hat{\phi}_{j,m_2}(Z_i, \hat{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0) \right)^2 = O_p(\xi_k^4 (n_1^{-2} + n_2^{-2})).
\]  

(195)

The claim in (185) follows from Assumption 4(iii), (186), (188), (191), (192), (193), (194) and (192).

(iv) By definition,

\[
n_j (\hat{\Omega}_{j,n} - \Omega_{j,n}) = \frac{(\hat{\phi}_{w,\theta,n} - \phi_{w,\theta,n}) (I_{d_g} \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) \hat{Q}_{n,j,u_j} (I_{d_g} \otimes Q_{n,j,k_j}^{-1} P'_{n,k_j}) \hat{\phi}_{w,\theta,n}}{n^2} \\
+ \frac{\phi_{w,\theta,n} (I_{d_g} \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) (\hat{Q}_{n,j,u_j} - Q_{n,j,u_j}) (I_{d_g} \otimes Q_{n,j,k_j}^{-1} P'_{n,k_j}) \hat{\phi}_{w,\theta,n}}{n^2} \\
+ \frac{\phi_{w,\theta,n} (I_{d_g} \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) Q_{n,j,u_j} (I_{d_g} \otimes Q_{n,j,k_j}^{-1} P'_{n,k_j}) (\hat{\phi}_{w,\theta,n} - \phi_{w,\theta,n})'}{n^2}.
\]  

(196)

By Lemma 4, Lemma 15 iii) and Assumption 4(iii),

\[(2C)^{-1} \leq \lambda_{\min} (\hat{Q}_{n,j,u_j}) \leq \lambda_{\max} (\hat{Q}_{n,j,u_j}) \leq 2C\]

(197)

with probability approaching 1. For \(l = 1, \ldots, d_{\theta}\), let \(\phi_{w,\theta_1,n}\) and \(\hat{\phi}_{w,\theta_1,n}\) denote the \(l\)-th rows of \(\phi_{w,\theta,n}\) and \(\hat{\phi}_{w,\theta,n}\) respectively. For any \(l = 1, \ldots, d_{\theta}\), by (43), (197) and Lemma 15(ii),

\[
\frac{(\hat{\phi}_{w,\theta_1,n} - \phi_{w,\theta_1,n}) (I_{d_g} \otimes P_{n,k_j} Q_{n,j,k_j}^{-1}) \hat{Q}_{n,j,u_j} (I_{d_g} \otimes Q_{n,j,k_j}^{-1} P'_{n,k_j}) (\hat{\phi}_{w,\theta_1,n} - \phi_{w,\theta_1,n})'}{n^2} \\
\leq \lambda_{\max} (\hat{Q}_{n,j,u_j}) \sum_{m=1}^{d_{\theta}} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n}) P_{n,k_j} Q_{n,j,k_j}^{-2} P'_{n,k_j} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n})' \\
= \lambda_{\max} (\hat{Q}_{n,j,u_j}) \sum_{m=1}^{d_{\theta}} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n}) P_{n,k_j} Q_{n,j,k_j}^{-2} P'_{n,k_j} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n})' \\
\leq \frac{\lambda_{\max} (\hat{Q}_{n,j,u_j})}{\lambda_{\min} (Q_{n,j,k_j})} \sum_{m=1}^{d_{\theta}} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n}) P_{n,k_j} P'_{n,k_j} (\hat{\phi}_{w,\theta_1,m,n} - \phi_{w,\theta_1,m,n})' \\
= O_p(n_1^{-1} + n_2^{-1} + \sigma_w^2),
\]

(198)
where \( \hat{\phi}_{w\theta_1,m,n} \) and \( \phi_{w\theta_1,m,n} \) denote the \( l \)-th rows of \( \phi_{w\theta_1,m,n} \) and \( \hat{\phi}_{w\theta_1,m,n} \) respectively. For any \( l_1, l_2 = 1, \ldots, d_\theta \), by Cauchy-Schwarz inequality and (198),

\[
\begin{align*}
\frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_2,n} - \phi_{w\theta_2,n})|}{n^2} \\
\leq \frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|^{1/2}}{n} \\
\times \frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|^{1/2}}{n} \\
= O_p(n_1^{-1} + n_2^{-1} + \delta_{w,n}^2). \tag{199}\end{align*}
\]

For any \( l = 1, \ldots, d_\theta \), by (43), (61) and (197),

\[
\begin{align*}
\frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}'_{w\theta_1,n})|}{n^2} \\
\leq \lambda_{\max}(\hat{Q}_{n_j,u_j})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1} P_{n,k_j}' \hat{\phi}'_{w\theta_1,n}) \\
= \lambda_{\max}(\hat{Q}_{n_j,u_j})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1} P_{n,k_j}' \hat{\phi}'_{w\theta_1,n}) \\
\leq \lambda_{\max}(\hat{Q}_{n_j,u_j}) \lambda_{\max}(Q_{n_j,k_j}) \sum_{m=1}^{d_\theta} \frac{\phi_{w\theta_1,n} P_{n,k_j} P_{n,k_j}' \phi_{w\theta_1,n}}{n^2} \\
\leq \lambda_{\max}(\hat{Q}_{n_j,u_j}) \lambda_{\max}(Q_{n_j,k_j}) \sum_{m=1}^{d_\theta} \frac{\phi_{w\theta_1,n} \phi_{w\theta_1,n}}{n} = O_p(1). \tag{200}\end{align*}
\]

By the Cauchy-Schwarz inequality, (198) and (200),

\[
\begin{align*}
\frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|}{n^2} \\
\leq \frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|^{1/2}}{n} \\
\times \frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|^{1/2}}{n} \\
= O_p(n_1^{-1/2} + n_2^{-1/2} + \delta_{w,n}). \tag{201}\end{align*}
\]

Collecting the results in (199) and (201), we get

\[
\frac{|(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})(I_{d_\theta} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})\hat{Q}_{n_j,u_j} (I_{d_\theta} \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}')(\hat{\phi}_{w\theta_1,n} - \phi_{w\theta_1,n})|}{n^2} = O_p(n_1^{-1/2} + n_2^{-1/2} + \delta_{w,n}). \tag{202}\]
For any $l = 1, \ldots, d_g$, by (43) and (61),

\[
\frac{\phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \phi'_{w\theta_1,n}}{n^2} = \sum_{m=1}^{d_g} \phi_{w\theta_1,m,n} P_{n,k_j} Q_{n_j,k_j}^{-2} P'_{n,k_j} \phi'_{w\theta_1,m,n} \frac{n}{n} \leq \frac{\lambda_{\text{max}}(Q_{n_j,k_j})}{(\lambda_{\text{min}}(Q_{n_j,k_j}))^2} \sum_{m=1}^{d_g} \phi_{w\theta_1,m,n} P_{n,k_j} (P'_{n,k_j} P_{n,k_j})^{-1} P'_{n,k_j} \phi'_{w\theta_1,m,n} \frac{n}{n} = O_p(1). \tag{203}
\]

By (43) and Lemma 15 ii),

\[
\frac{\phi_{w\theta_1,n} - \phi_{w\theta_1,n}}{(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) (\phi_{w\theta_1,n} - \phi_{w\theta_1,n})'} \frac{n}{n} = \sum_{m=1}^{d_g} \phi_{w\theta_1,m,n} P_{n,k_j} Q_{n_j,k_j}^{-2} P'_{n,k_j} (\phi_{w\theta_1,m,n} - \phi_{w\theta_1,m,n})' \frac{n}{n} \leq \frac{1}{(\lambda_{\text{min}}(Q_{n_j,k_j}))^2} \sum_{m=1}^{d_g} \phi_{w\theta_1,m,n} P_{n,k_j} P'_{n,k_j} (\phi_{w\theta_1,m,n} - \phi_{w\theta_1,m,n})' \frac{n}{n} = O_p(n_1^{-1} + n_2^{-1} + \delta_{w,n}). \tag{204}
\]

For any $l = 1, \ldots, d_g$, by the Cauchy-Schwarz inequality, Lemma 15 iii) and (203),

\[
\frac{\phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(\hat{Q}_{n_j,u_j} - Q_{n_j,u_j})^2 (I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \phi'_{w\theta_1,n}}{n^2} \leq \left\| \hat{Q}_{n_j,u_j} - Q_{n_j,u_j} \right\| \frac{2 \phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \phi'_{w\theta_1,n}}{n^2} = O_p(\epsilon_{k_j}^2 l_j (n_1^{-1} + n_2^{-1})). \tag{205}
\]

By the triangle inequality, Cauchy-Schwarz inequality, (203), (204) and (205),

\[
\frac{\phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(\hat{Q}_{n_j,u_j} - Q_{n_j,u_j})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \phi'_{w\theta_1,n}}{n^2} \leq \left| \phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(\hat{Q}_{n_j,u_j} - Q_{n_j,u_j})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \phi'_{w\theta_1,n} \right| \frac{n}{n} + \left| \phi_{w\theta_1,n}(I_{d_g} \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(\hat{Q}_{n_j,u_j} - Q_{n_j,u_j})(I_{d_g} \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) (\phi_{w\theta_1,n} - \phi_{w\theta_1,n}) \right| \frac{n}{n} = O_p(\epsilon_{k_j} l_j^{1/2}) (n_1^{-1/2} + n_2^{-1/2})), \tag{206}
\]
for any \( l_1, l_2 = 1, \ldots, d_\theta \), which implies that

\[
\frac{\phi_{w,\theta,n}(I_d \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(Q_{n_j,u_j} - Q_{n_j,u_j})(I_d \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j})}{n^2} \phi'_{w,\theta,n} = O_p(\xi_{k_j}^{1/2} (n_1^{-1/2} + n_2^{-1/2})).
\]

By Cauchy-Schwarz inequality, Lemma 4, (203) and (204),

\[
\left| \frac{\phi_{w,\theta_1,n}(I_d \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})Q_{n_j,u_j}(I_d \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j})}{n^4} \left( \phi'_{w,\theta_1,n} - \phi'_{w,\theta_2,n} \right) \right|^2
\leq \left( \lambda_{\max}(Q_{n_j,u_j}) \right)^2 \frac{\phi_{w,\theta_1,n}(I_d \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})(I_d \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j})}{n^2} \phi'_{w,\theta_1,n}
\times \left( \phi'_{w,\theta_2,n} - \phi'_{w,\theta_2,n} \right)
= O_p(n_1^{-1} + n_2^{-1} + \delta_{w,n}),
\]

which implies that

\[
\frac{\phi_{w,\theta,n}(I_d \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})}{n^2} Q_{n_j,u_j}(I_d \otimes Q_{n_j,k_j}^{-1} P'_{n,k_j}) \left( \phi_{w,\theta,n} - \phi_{w,\theta,n} \right) = O_p(n_1^{-1/2} + n_2^{-1/2} + \delta_{w,n}).
\]

Collecting the results in (196), (202), (207) and (209), we get

\[
n_j(\hat{\Omega}_{j,n} - \Omega_{j,n}) = O_p(\xi_{k_j}^{1/2} (n_1^{-1/2} + n_2^{-1/2}) + \delta_{w,n}) = o_p(1),
\]

where the second equality is by Assumption 4(iii), which proves the claim.

(v) By Assumptions 2(ii) and 4(iii), Lemma 15(i) and (143),

\[
(2C)^{-1} \leq \lambda_{\min}(\hat{H}_n) \leq \lambda_{\max}(\hat{H}_n) \leq 4C
\]
with probability approaching 1. By Lemma 4, Lemma 8 and (43),

\[ n_j \Omega_{j,n} = \frac{\phi_{w\theta,n}(I_m \otimes P_{n,k_j} Q_{n_j,k_j}^{-1})Q_{n_j,u_j}(I_m \otimes Q_{n_j,k_j}^{-1} P_{n,k_j}' \phi_{w\theta,n})}{n^2} \]

\[ \leq \lambda_{\max}(Q_{n_j,u_j}) \frac{\phi_{w\theta,n}(I_m \otimes P_{n,k_j} Q_{n_j,k_j}^{-2} P_{n,k_j}' \phi_{w\theta,n})}{n^2} \]

\[ = \lambda_{\max}(Q_{n_j,u_j}) \sum_{m=1}^{d_q} \phi_{w\theta,m,n} P_{n,k_j} Q_{n_j,k_j}^{-2} P_{n,k_j}' \phi_{w\theta,m,n} \]

\[ \leq \lambda_{\max}(Q_{n_j,u_j}) \lambda_{\max}(Q_{n,k_j}) \frac{d_q}{\lambda_{\min}(Q_{n_j,k}_j)} \sum_{m=1}^{d_q} n^{-1} \phi_{w\theta,m,n} P_{n,k_j} (P_{n,k_j}' P_{n,k_j}) P_{n,k_j}' \phi_{w\theta,m,n} \]

\[ = \lambda_{\max}(Q_{n_j,u_j}) \frac{\lambda_{\max}(Q_{n,k_j})}{\lambda_{\min}(Q_{n_j,k}_j)} \sum_{m=1}^{d_q} n^{-1} \phi_{w\theta,m,n} P_{n,k_j} (P_{n,k_j}' P_{n,k_j}) P_{n,k_j}' \phi_{w\theta,m,n} \]

\[ \leq \lambda_{\max}(Q_{n_j,u_j}) \lambda_{\max}(Q_{n,k_j}) \frac{d_q}{\lambda_{\min}(Q_{n_j,k}_j)} \sum_{m=1}^{d_q} n^{-1} \phi_{w\theta,m,n} P_{n,k_j} (P_{n,k_j}' P_{n,k_j}) P_{n,k_j}' \phi_{w\theta,m,n} \]

which together with Assumption (iv), Lemma 4 (43) and (58) implies that

\[ \lambda_{\max}(n_j \Omega_{j,n}) \leq C \]

(212) with probability approaching 1. By (131), (210) and (212),

\[ (2C)^{-1} \leq \lambda_{\min}(n_j \Omega_{j,n}) \leq \lambda_{\max}(n_j \Omega_{j,n}) \leq 2C \]

(213) with probability approaching 1. By the triangle inequality

\[ \left\| n_j^{-1}(\hat{H}_n (\Omega_{1,n} + \Omega_{2,n})^{-1} \hat{H}_n - H_n (\Omega_{1,n} + \Omega_{2,n})^{-1} H_n) \right\| \]

\[ \leq \left\| n_j^{-1}(\hat{H}_n - H_n) (\Omega_{1,n} + \Omega_{2,n})^{-1} \hat{H}_n \right\| \]

\[ + \left\| n_j^{-1} H_n ((\Omega_{1,n} + \Omega_{2,n})^{-1} - (\Omega_{1,n} + \Omega_{2,n})^{-1}) \hat{H}_n \right\| \]

\[ + \left\| n_j^{-1} H_n (\Omega_{1,n} + \Omega_{2,n})^{-1} (\hat{H}_n - H_n) \right\|. \]

(214) By Lemma 15(i), (211) and (213),

\[ \left\| n_j^{-1}(\hat{H}_n - H_n) (\Omega_{1,n} + \Omega_{2,n})^{-1} \hat{H}_n \right\| \]

\[ \leq \left\| \hat{H}_n - H_n \right\| \left\| n_j^{-1}(\Omega_{1,n} + \Omega_{2,n})^{-1} \hat{H}_n \right\| \]

\[ = \left\| \hat{H}_n - H_n \right\| \left( \text{tr} \left( n_j^{-2}(\Omega_{1,n} + \Omega_{2,n})^{-1} \hat{H}_n^2 (\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right)^{1/2} \]

\[ \leq \frac{d_q \lambda_{\max}(\hat{H}_n)}{\lambda_{\min}(n_j \Omega_{j,n})} \left\| \hat{H}_n - H_n \right\| = o_p(1). \]

(215)
By Assumption 3(ii), (131), (143) and Lemma 15(i),

\[
\left\| n_j^{-1}H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}(\bar{H}_n - H_n) \right\| \\
\leq \left\| n_j^{-1}H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right\| \left\| \bar{H}_n - H_n \right\| \\
= \left\| \bar{H}_n - H_n \right\| \left( \text{tr} \left( n_j^{-2}H_n(\Omega_{1,n} + \Omega_{2,n})^{-2} \right) H_n \right)^{1/2} \\
\leq \frac{d\lambda_{\max}(H_n)}{\lambda_{\min}(n_j\Omega_{j,n})} \left\| \bar{H}_n - H_n \right\| = o_p(1). 
\] 

(216)

By the triangle inequality, Assumption 4(iii), (143), (209), (210) and (213),

\[
\left\| n_j^{-1}H_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - (\Omega_{1,n} + \Omega_{2,n})^{-1}\hat{H}_n \right\| \\
\leq \left\| n_j^{-1}H_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1}(\hat{\Omega}_{1,n} - \Omega_{1,n})(\Omega_{1,n} + \Omega_{2,n})^{-1}\hat{H}_n \right\| \\
+ \left\| n_j^{-1}H_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1}(\hat{\Omega}_{2,n} - \Omega_{2,n})(\Omega_{1,n} + \Omega_{2,n})^{-1}\hat{H}_n \right\| \\
\leq \left\| n_j^{-1}H_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} \right\| \left\| n_1^{-1}(\Omega_{1,n} + \Omega_{2,n})^{-1}\hat{H}_n \right\| \left\| n_2(\hat{\Omega}_{1,n} - \Omega_{1,n}) \right\| \\
+ \left\| n_j^{-1}H_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} \right\| \left\| n_2^{-1}(\Omega_{1,n} + \Omega_{2,n})^{-1}\hat{H}_n \right\| \left\| n_2(\hat{\Omega}_{2,n} - \Omega_{2,n}) \right\| \\
= \frac{d\lambda_{\max}(\hat{H}_n)\lambda_{\max}(H_n)}{\lambda_{\min}(n_j\Omega_{j,n})} \left( \frac{\left\| n_1(\hat{\Omega}_{1,n} - \Omega_{1,n}) \right\|}{\lambda_{\min}(n_1\Omega_{1,n})} + \frac{\left\| n_2(\hat{\Omega}_{2,n} - \Omega_{2,n}) \right\|}{\lambda_{\min}(n_2\Omega_{2,n})} \right) \\
= O_p(\max_{j=1,2} \lambda_{k,j}^{1/2}(n_1^{-1/2} + n_2^{-1/2})) = o_p(1) 
\] 

(217)

Combining the results in (214), (215), (216) and (217), we get

\[
\left\| n_j^{-1}(\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1}H_n) \right\| = o_p(1). 
\] 

(218)

By Assumption 3(ii), (212) and (218),

\[
\left\| (\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| \\
= \left\| (\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| \\
\leq \left\| (\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| \\
+ \left\| (\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| \\
\leq \frac{\lambda_{\max}(n_1\Omega_{1,n})}{\lambda_{\min}(H_n)^2} \left\| n_1^{-1}(\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| \\
+ \frac{\lambda_{\max}(n_2\Omega_{2,n})}{\lambda_{\min}(H_n)^2} \left\| n_2^{-1}(\hat{H}_n(\hat{\Omega}_{1,n} + \hat{\Omega}_{2,n})^{-1} - H_n(\Omega_{1,n} + \Omega_{2,n})^{-1})H_n \right\| = o_p(1) 
\] 

(219)

which finishes the proof. □
Proof of Theorem 4. The claim in (21) follows from Lemma 15(v). We next prove the claim in (22). For any \( \gamma_n \in \mathbb{R}^d \) with \( \gamma_n^T \gamma_n = 1 \),

\[
\left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| = \left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| \\
= \left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| \\
\leq \left( \frac{\lambda_{\max}(n_j \Omega_{j,n})}{n_j^{1/2} \lambda_{\min}(H_n)} \right)^{1/2} \left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| \\
\leq \left( \frac{\lambda_{\max}(n_j \Omega_{j,n})}{n_j^{1/2} \lambda_{\min}(H_n)} \right)^{1/2} \left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| \\
\leq \left( \frac{\lambda_{\max}(n_j \Omega_{j,n})}{\lambda_{\min}(n_j \Omega_{j,n})} \right)^{1/2} \left\| \gamma_n^T \left( \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - I_d \right) \right\| = o_p(1),
\]

(220)

where the second equality is by Assumption 2(ii), (131), (143), (212) and Lemma 15(v). By the Cauchy-Schwarz inequality, (10) and (220),

\[
\left| \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right| \leq \left\| \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right\| \\
= \left\| \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right\| \\
\leq \left\| \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) - \gamma_n^T \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right\| \\
\times \left\| \left( H_n(\Omega_{1,n} + \Omega_{2,n})^{-1} \right) \right\| = o_p(1).
\]

(221)

The claim in (21) follows from Theorem 2, (221) and the Slutsky theorem. \( \blacksquare \)

Lemma 16. Under Assumptions 2, 3, 4 and 5, we have

\[
\sup_{z \in \mathcal{Z}} \left| \hat{\Sigma}_j(z) - \Sigma_j(z) \right| = O_p(\xi_{kj} k_j^{-\alpha} + \xi_{kj} k_j^{1/2} n_j^{-1/2} + \xi_{kj} (n_1^{-1/2} + n_2^{-1/2})).
\]

Proof of Lemma 16. By definition,

\[
\tilde{u}_{j,m,i} = u_{j,m,i} + \left[ g_{j,m}(Y_{j,i}; \theta_n) - g_{j,m}(Y_{j,i}; \theta_0) \right] - \left[ \phi_{j,m}(Z_i; \tilde{\theta}_n) - \phi_{j,m}(Z_i; \theta_0) \right],
\]

(222)

where \( \tilde{\theta}_n \) is the SMD estimator with the identity weight function. We first show that

\[
\sup_{z \in \mathcal{Z}} \left| P_{kj}^*(z) \left( P_{nj,kj}^* P_{n,j,kj} \right)^{-1} P_{nj,kj}^* U_{j,m_1,m_2,n_j} - \Sigma_{j,m_1,m_2}(z) \right| = O_p(\xi_{kj} (k_j^{1/2} n_j^{-1/2} + k_j^{-\alpha})),
\]

(223)
where $U_{j,m_1,m_2,n_j} = ((u_{j,m_1,i}u_{j,m_2,i})_{i \in I_j})'$. Let $\Sigma_{j,m_1,m_2,k_j}(z) = P_{k_j}(z)\beta_{\Sigma_{j,m_1,m_2,k_j}}$. Then

\[
(P_{n_j,k_j}P_{n_j,k_j})^{-1}P_{n_j,k_j}U_{j,m_1,m_2,n_j} - \beta_{\Sigma_{j,m_1,m_2,k_j}}
\]

\[
= (P_{n_j,k_j}P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_{k_j}(Z_i)(u_{j,m_1,i}u_{j,m_2,i} - \Sigma_{j,m_1,m_2}(Z_i))
\]

\[
+ (P_{n_j,k_j}P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_{k_j}(Z_i)(\Sigma_{j,m_1,m_2}(Z_i) - P_{k_j}(Z_i)\beta_{\Sigma_{j,m_1,m_2,k_j}}).
\]

(224)

By Assumptions 1(i), 2(v) and (43),

\[
E \left\| (P_{n_j,k_j}P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_{k_j}(Z_i)(u_{j,m_1,i}u_{j,m_2,i} - \Sigma_{j,m_1,m_2}(Z_i)) \right\|^2_{\{Z_i\} \in I}
\]

\[
\leq \sup_{z \in Z} E[u_{j,m_1}^2u_{j,m_2}^2|Z = z]tr((P_{n_j,k_j}P_{n_j,k_j})^{-1}) = O_p(kj_{n_j})
\]

(225)

which together with the Markov inequality implies that

\[
(P_{n_j,k_j}P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_{k_j}(Z_i)(u_{j,m_1,i}u_{j,m_2,i} - \Sigma_{j,m_1,m_2}(Z_i)) = O_p(k_j^{1/2}n_j^{-1/2}).
\]

(226)

By Assumption 1(ii) and (43),

\[
\left\| (P_{n_j,k_j}P_{n_j,k_j})^{-1}\sum_{i \in I_j} P_{k_j}(Z_i)(\Sigma_{j,m_1,m_2}(Z_i) - P_{k_j}(Z_i)\beta_{\Sigma_{j,m_1,m_2,k_j}}) \right\|^2
\]

\[
\leq (\lambda_{\min}(Q_{n_j,k_j}))^{-1}n_j^{-1}\sum_{i \in I_j} (\Sigma_{j,m_1,m_2}(Z_i) - P_{k_j}(Z_i)\beta_{\Sigma_{j,m_1,m_2,k_j}})^2 = O_p(k_j^{-2r_u}).
\]

(227)

By the triangle inequality, the Cauchy-Schwarz inequality Assumption 4(ii), (224), (226) and (227),

\[
\sup_{z \in Z} \left| P_{k_j}'(z)(P_{n_j,k_j}P_{n_j,k_j})^{-1}P_{n_j,k_j}U_{j,m_1,m_2,n_j} - \Sigma_{j,m_1,m_2}(z) \right|
\]

\[
\leq \sup_{z \in Z} \left| P_{k_j}'(z) \left[ (P_{n_j,k_j}P_{n_j,k_j})^{-1}P_{n_j,k_j}U_{j,m_1,m_2,n_j} - \beta_{\Sigma_{j,m_1,m_2,k_j}} \right] \right|
\]

\[
+ \sup_{z \in Z} \left| P_{k_j}'(z)\beta_{\Sigma_{j,m_1,m_2,k_j}} - \Sigma_{j,m_1,m_2}(z) \right|
\]

\[
\leq \xi_{k_j} \left\| (P_{n_j,k_j}P_{n_j,k_j})^{-1}P_{n_j,k_j}U_{j,m_1,m_2,n_j} - \beta_{\Sigma_{j,m_1,m_2,k_j}} \right\| + O(k_j^{-r_u})
\]

\[
= O_p(\xi_{k_j}(k_j^{1/2}n_j^{-1/2} + k_j^{-r_u})),
\]

(228)

which proves (223). Since the identity weight function satisfies Assumption 1(iv), we can invoke Theorem 2 to deduce that

\[
\tilde{\theta}_n - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2}).
\]

(229)
By the Cauchy-Schwarz inequality, Assumption 4(i), (172) and (229),

\[ \sup_{z \in Z} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z) Q_{n_j, k_j} P_{k_j}(Z_i) u_{j,m_1,i}(g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) \right| \]

\[ \leq \sup_{z \in Z} \left| P_{k_j}(z) Q_{n_j, k_j} \right| n_j^{-1} \left| \sum_{i \in I_j} P_{k_j}(Z_i) u_{j,m_1,i}(g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) \right| \]

\[ \leq \xi_k \lambda_{\max}(Q_{n_j, k_j}) \left( \sum_{i \in I_j} u_{j,m_1,i}^2 (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^2 \right)^{1/2} \]

\[ \leq \xi_k \lambda_{\max}(Q_{n_j, k_j}) \left( \sum_{i \in I_j} u_{j,m_1,i}^4 \right)^{1/4} \left( \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^4 \right)^{1/4} \]

(230)

By the Cauchy-Schwarz inequality, Assumption 4(i), (172) and (229),

\[ n_j^{-1} \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^4 \]

\[ \leq \left\| \tilde{\theta}_n - \theta_0 \right\|_4^4 \sup_{\theta \in \Theta} n_j^{-1} \sum_{i \in I_j} \left\| g_{j,m}^{(1)}(Y_{j,i}, \theta) \right\|_4^4 = O_p(n_1^{-2} + n_2^{-2}). \]

(231)

By Assumption 2(v) and the Markov inequality, \( n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i}^4 = O_p(1) \) which together with (43), (230) and (231) implies that

\[ \sup_{z \in Z} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z) Q_{n_j, k_j} P_{k_j}(Z_i) u_{j,m_1,i}(g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) \right| = O_p(\xi_k (n_1^{-1/2} + n_2^{-1/2})) \]

(232)

for any \( m_1, m_2 = 1, \ldots, d_g \). By the Cauchy-Schwarz inequality,

\[ \sup_{z \in Z} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z) Q_{n_j, k_j} P_{k_j}(Z_i) u_{j,m_1,i}(\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0)) \right| \]

\[ \leq \sup_{z \in Z} \left| P_{k_j}(z) Q_{n_j, k_j} \right| n_j^{-1} \left| \sum_{i \in I_j} P_{k_j}(Z_i) u_{j,m_1,i}(\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0)) \right| \]

\[ \leq \xi_k \lambda_{\max}(Q_{n_j, k_j}) \left( \sum_{i \in I_j} u_{j,m_1,i}^2 (\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0))^2 \right)^{1/2} \]

(233)

By Assumptions 2(v), 4(iii), (43) and (188),

\[ \lambda_{\max}(Q_{n_j, u_{j,m_1}}) \leq \sup_{z \in Z} E[u_{j,m_1}^2 | Z = z] \lambda_{\max}(Q_{n_j, k_j}) \leq 2C \]

(234)

68
with probability approaching 1. By definition, 
\[ \hat{\phi}_{j,m_2}(Z_i, \theta) = P_{k_j}(Z_i)'(P'_{n_j,k_j} P_{n_j,k_j})^{-1}P'_{n_j,k_j} g_{j,m,n_j}(\theta) \]
Then by (43), (234) and similar arguments in showing (231),
\[
\begin{align*}
n_j^{-1} & \sum_{i \in I_j} u_{j,m_1,i}^{2}(\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0))^2 \\
& = \left( g_{j,m_2,n_j}(\tilde{\theta}_n) - g_{j,m_2,n_j}(\theta_0) \right)' P_{n_j,k_j} (P'_{n_j,k_j} P_{n_j,k_j})^{-1} \\
& \times Q_{n_j,n_j,m_1}(P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} \left( g_{j,m_2,n_j}(\tilde{\theta}_n) - g_{j,m_2,n_j}(\theta_0) \right) \\
& \leq \frac{\lambda_{\text{max}}(Q_{n_j,n_j,m_1})}{n_j \lambda_{\text{min}}(Q_{n_j,k_j})} n_j^{-1} \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}; \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}; \theta_0))^2 = O_p(n_1^{-1} + n_2^{-1}),
\end{align*}
\] (235)
which together with (43) and (233) implies that
\[
\sup_{z \in Z} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)' Q_{n_j,k_j} P_{k_j}(Z_i) u_{j,m_1,i}(\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \hat{\phi}_{j,m_2}(Z_i, \theta_0)) \right| = O_p(\xi_{k_j}(n_1^{-1/2} + n_2^{-1/2})).
\] (236)
By the Cauchy-Schwarz inequality,
\[
\begin{align*}
\sup_{z \in Z} & \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)' Q_{n_j,k_j} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0)) \right| \\
& \leq \sup_{z \in Z} \| P_{k_j}(z) \| \left| n_j^{-1} Q_{n_j,k_j} \sum_{i \in I_j} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0)) \right| .
\end{align*}
\] (237)
By Assumptions 2(v) and 3(i), and (43),
\[
\begin{align*}
E & \left[ \left| n_j^{-1} Q_{n_j,k_j} \sum_{i \in I_j} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0)) \right|^2 \right]_{\{Z_i\} \in I} \\
& = tr \left( n_j^{-2} Q_{n_j,k_j} \sum_{i \in I_j} E[u_{j,m_1,i}^2 | Z_i] (\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0))^2 P_{k_j}(Z_i) P_{k_j}(Z_i)' Q_{n_j,k_j} \right) \\
& \leq \sup_{z \in Z} E \left[ u_{j,m_1}^2 | Z = z \right] (\lambda_{\text{max}}(Q_{n_j,k_j}))^2 n_j \sum_{i \in I_j} (\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0))^2 \| P_{k_j}(Z_i) \|^2 \\
& \leq \sup_{z \in Z} E \left[ u_{j,m_1}^2 | Z = z \right] (\lambda_{\text{max}}(Q_{n_j,k_j}))^2 n_j^2 \sum_{i \in I_j} (\phi_{j,m,k_j}(Z_i, \theta_0) - \phi_{j,m}(Z_i, \theta_0))^2 \\
& = O_p(\xi_{k_j}^2 n_j^{-1} k_j^{-2r^2}).
\end{align*}
\] (238)
where \( \phi_{j,m,k}(z, \theta_0) = P_k(z)\beta_{j,m,k} \), which together with (237) implies that

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z) Q_{n_j,k} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k}(Z_i, \theta_0) - \phi_{j,m,k}(Z_i, \theta_0)) \right| = O_P(\xi_j^2 k_j n_j^{-1/2}).
\]

By the Cauchy-Schwarz inequality,

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z) Q_{n_j,k} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k}(Z_i, \theta_0) - \phi_{j,m,k}(Z_i, \theta_0)) \right|
\leq \sup_{z \in \mathbb{Z}} \left| P_{k_j}(z) Q_{n_j,k} \right| \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(Z_i) u_{j,m_1,i}(\phi_{j,m,k}(Z_i, \theta_0) - \phi_{j,m,k}(Z_i, \theta_0)) \right|
\leq \xi_j \lambda_{\max}(Q_{n_j,k}) \left| n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i} P_{k_j}(Z_i) P_{k_j}(Z_i) \right|
\times \left| (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} g_{j,m,n_j}(\theta_0) - \beta_{j,m,k} \right|.
\]

By Assumptions 1(i), 2(v) and (43),

\[
E \left[ \left| n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i} P_{k_j}(Z_i) P_{k_j}(Z_i) \right| \{Z_i \in \mathcal{I} \} \right]^{1/2} \]
\[
= \text{tr} \left( n_j^{-2} \sum_{i \in I_j} E \left[ u_{j,m_1,i}^2 | Z_i \right] \left( P_{k_j}(Z_i) P_{k_j}(Z_i) \right)^2 \right)
\leq n_j^{-1} \sup_{z \in \mathbb{Z}} E \left[ u_{j,m_1,i}^2 | Z = z \right] \sup_{i \in \mathbb{Z}} \left| P_{k_j}(z) \right|^2 \text{tr}(Q_{n_j,k}) = O_P(\xi_j^2 k_j n_j^{-1})
\]

which together with the Markov inequality implies that

\[
\left| n_j^{-1} \sum_{i \in I_j} u_{j,m_1,i} P_{k_j}(Z_i) P_{k_j}(Z_i) \right| = O_P(\xi_j^{1/2} k_j^{1/2} n_j^{-1/2}).
\]

By the triangle inequality, (178) and (179),

\[
\left| (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} g_{j,m,n_j}(\theta_0) - \beta_{j,m,k} \right|
\leq \left| (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} (g_{j,m,n_j}(\theta_0) - \phi_{j,m,n_j}(\theta_0)) \right|
+ \left| (P'_{n_j,k_j} P_{n_j,k_j})^{-1} P'_{n_j,k_j} (\phi_{j,m,n_j}(\theta_0) - \phi_{j,m,k,j}(\theta_0)) \right|
= O_P(\xi_j^{1/2} k_j n_j^{-1/2} + k_j^{-r_\phi})
\]
which together with (240) and (242) implies that

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)^T Q_{n_j, k_j} P_{k_j}(Z_i) u_{j,m_1,i}(\hat{\phi}_{j,m_2}(Z_i, \theta_0) - \phi_{j,m_2,k_j}(Z_i, \theta_0)) \right| \\
= O_p(\xi_{k_j}^2 k_j^{1/2} n_j^{-1/2} (k_j^{-1/2} n_j^{-1/2} + k_j^{-r^*})).
\]  

(244)

By the triangle inequality, Assumption III(iii), (236), (239) and (244),

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)^T Q_{n_j, k_j} P_{k_j}(Z_i) u_{j,m_1,i}(\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \phi_{j,m_2,k_j}(Z_i, \theta_0)) \right| \\
= O_p(\xi_{k_j} (n_1^{-1/2} + n_2^{-1/2}) + \xi_{k_j}^2 k_j n_j^{-1}).
\]  

(245)

By the triangle inequality and Cauchy-Schwarz inequality, (43) and (173),

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)^T Q_{n_j, k_j} P_{k_j}(Z_i) (g_{j,m_1}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0)) (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) \right| \\
\leq \sup_{z \in \mathbb{Z}} \left\| P_{k_j}(z)^T Q_{n_j, k_j}^{1/2} \right\|^2 \left| n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0)) (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0)) \right| \\
\leq \xi_{k_j}^2 \lambda_{\max}(Q_{n_j, k_j}) \left( n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0))^2 \right)^{1/2} \\
\times \left( n_j^{-1} \sum_{i \in I_j} (g_{j,m_2}(Y_{j,i}, \tilde{\theta}_n) - g_{j,m_2}(Y_{j,i}, \theta_0))^2 \right)^{1/2} = O_p(\xi_{k_j}^2 (n_1^{-1} + n_2^{-1})).
\]  

(246)

By the triangle inequality and the Cauchy-Schwarz inequality, (43) and (170),

\[
\sup_{z \in \mathbb{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)^T Q_{n_j, k_j} P_{k_j}(Z_i) (\hat{\phi}_{j,m_1}(Z_i, \tilde{\theta}_n) - \phi_{j,m_1}(Z_i, \theta_0)) (\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0)) \right| \\
\leq \sup_{z \in \mathbb{Z}} \left\| P_{k_j}(z)^T Q_{n_j, k_j}^{1/2} \right\|^2 \left| n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_1}(Z_i, \tilde{\theta}_n) - \phi_{j,m_1}(Z_i, \theta_0)) (\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0)) \right| \\
\leq \xi_{k_j}^2 \lambda_{\max}(Q_{n_j, k_j}) \left( n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_1}(Z_i, \tilde{\theta}_n) - \phi_{j,m_1}(Z_i, \theta_0))^2 \right)^{1/2} \\
\times \left( n_j^{-1} \sum_{i \in I_j} (\hat{\phi}_{j,m_2}(Z_i, \tilde{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0))^2 \right)^{1/2} = O_p(\xi_{k_j}^2 (n_1^{-1} + n_2^{-1} + k_j n_j^{-1} + k_j^{-2r^*})).
\]  

(247)
By the triangle inequality and the Cauchy-Schwarz inequality, (43), (170) and (173),

\[
\sup_{z \in \mathcal{Z}} \left| n_j^{-1} \sum_{i \in I_j} P_{k_j}(z)^i Q_{n_j,k_j} P_{k_j}(Z_i)^i (g_{j,m_1}(Y_{j,i}, \overline{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0)) (\widehat{\phi}_{j,m_2}(Z_i, \overline{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0)) \right|
\leq \sup_{z \in \mathcal{Z}} \left\| P_{k_j}(z)^i Q_{n_j,k_j}^{1/2} \right\|^2 n_j^{-1} \sum_{i \in I_j} \left| (g_{j,m_1}(Y_{j,i}, \overline{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0)) (\widehat{\phi}_{j,m_2}(Z_i, \overline{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0)) \right|
\leq \xi_{k_j}^2 \lambda_{\text{max}}(Q_{n_j,k_j}) \left( n_j^{-1} \sum_{i \in I_j} (g_{j,m_1}(Y_{j,i}, \overline{\theta}_n) - g_{j,m_1}(Y_{j,i}, \theta_0))^2 \right)^{1/2}
\times \left( n_j^{-1} \sum_{i \in I_j} (\widehat{\phi}_{j,m_2}(Z_i, \overline{\theta}_n) - \phi_{j,m_2}(Z_i, \theta_0))^2 \right)^{1/2}
= O_p(\xi_{k_j}^2 (n_1^{-1} + n_2^{-1} + k_j n_j^{-1} + k_j^{-2r})).
\]

Applying Assumption (iii), and combining the results in (222), (228), (232), (245), (246), (247) and (248), we get

\[
\sup_{z \in \mathcal{Z}} \left| \Sigma_{j,m_1,m_2}(z) - \Sigma_{j,m_1,m_2}(z) \right| = O_p(\xi_{k_j} k_j^{-r} + \xi_{k_j} k_j^{1/2} n_j^{-1/2} + \xi_{k_j} (n_1^{-1/2} + n_2^{-1/2}))
\]

which finishes the proof. \(\blacksquare\)

**Proof of Theorem 5.** By Lemma 16 Assumptions (v) and (iii),

\[
\inf_{z \in \mathcal{Z}} \lambda_{\text{min}}(\Sigma_j(z)) \geq (2C)^{-1}
\]

with probability approaching 1. Therefore,

\[
\left\| (n_1^{-1} \Sigma_1(z) + n_2^{-1} \Sigma_2(z))^{-1} \right\| \leq d_\theta(\lambda_{\text{min}}(n_1^{-1} \Sigma_1(z) + n_2^{-1} \Sigma_2(z)))^{-1}
\leq d_\theta(n_1^{-1} \lambda_{\text{min}}(\Sigma_1(z)) + n_2^{-1} \lambda_{\text{min}}(\Sigma_2(z)))^{-1}
\leq 2C d_\theta(n_1^{-1} + n_2^{-1})^{-1}
\]

with probability approaching 1. Similarly

\[
\left\| (n_1^{-1} \Sigma_1(z) + n_2^{-1} \Sigma_2(z))^{-1} \right\| \leq d_\theta(\lambda_{\text{min}}(n_1^{-1} \Sigma_1(z) + n_2^{-1} \Sigma_2(z)))^{-1}
\leq d_\theta(n_1^{-1} \lambda_{\text{min}}(\Sigma_1(z)) + n_2^{-1} \lambda_{\text{min}}(\Sigma_2(z)))^{-1}
\leq 2C d_\theta(n_1^{-1} + n_2^{-1})^{-1}.
\]

72
By the triangle inequality, (250) and (251),

\[
\sup_{z \in \mathcal{Z}} \left\| \hat{W}_n^*(z) - W^*(z) \right\| \\
= \left( n_1^{-1} + n_2^{-1} \right) \sup_{z \in \mathcal{Z}} \left\| \left( n_1^{-1} \hat{\Sigma}_1(z) + n_2^{-1} \hat{\Sigma}_2(z) \right)^{-1} - \left( n_1^{-1} \Sigma_1(z) + n_2^{-1} \Sigma_2(z) \right)^{-1} \right\|
\leq \left( n_1^{-1} + n_2^{-1} \right) \sup_{z \in \mathcal{Z}} \left\| \left( \frac{\hat{\Sigma}_1(z) + \hat{\Sigma}_2(z)}{n_1} \right)^{-1} - \left( \frac{\hat{\Sigma}_1(z) - \Sigma_1(z)}{n_1} \right) \right\| \left\| \left( \frac{\hat{\Sigma}_1(z) + \Sigma_1(z)}{n_1} \right)^{-1} \right\|
\]
\[
+ \left( n_1^{-1} + n_2^{-1} \right) \sup_{z \in \mathcal{Z}} \left\| \left( \frac{\hat{\Sigma}_1(z) + \hat{\Sigma}_2(z)}{n_2} \right)^{-1} - \left( \frac{\Sigma_2(z) - \hat{\Sigma}_2(z)}{n_2} \right) \right\| \left\| \left( \frac{\Sigma_1(z) + \hat{\Sigma}_2(z)}{n_2} \right)^{-1} \right\|
\leq 4C^2 d_0^2 (n_1^{-1} + n_2^{-1})^{-1} \sup_{z \in \mathcal{Z}} \left[ \left\| \frac{\hat{\Sigma}_1(z) - \Sigma_1(z)}{n_1} \right\| + \left\| \frac{\hat{\Sigma}_2(z) - \Sigma_2(z)}{n_2} \right\| \right] \tag{252}
\]

with probability approaching 1, which together with Lemma 16 implies that

\[
\sup_{z \in \mathcal{Z}} \left\| \hat{W}_n^*(z) - W^*(z) \right\| \\
= 4C^2 d_0^2 (n_1^{-1} + n_2^{-1})^{-1} Op(\left( n_1^{-1} + n_2^{-1} \right) \max_{j=1,2} \left( \xi_{k_j} \left( k_j^{-r_u} + k_j^{1/2} n_j^{-1/2} + n_1^{-1/2} + n_2^{-1/2} \right) \right))
= Op(\max_{j=1,2} \left( \xi_{k_j} \left( k_j^{-r_u} + k_j^{1/2} n_j^{-1/2} + n_1^{-1/2} + n_2^{-1/2} \right) \right)). \tag{253}
\]

This finishes the proof. \( \blacksquare \)