SUPPLEMENT TO "Inference in Additively Separable Models with a High-Dimensional Set of Conditioning Variables"

Damian Kozbur
University of Zürich
Department of Economics
Schönberggasse 1, 8001 Zürich
email: `damian.kozbur@econ.uzh.ch`

Appendix A. Implementation Details

A.1. **Lasso Implementation Details.**

A.1.1. *Lasso implementation given penalty λ.* In every case, penalty loadings ℓ_j are chosen as described in [1] with one small modification. The procedure suggested in [1] requires an initial penalty loadings which are constructed using initial estimates of regression residuals. Their suggestion is to use $\widehat{\varepsilon}_i^{\text{initial}} = y_i$ followed by an iterative procedure. Here, instead, $\widehat{\varepsilon}_i^{\text{initial}}$ are taken as the linear regression residuals after regressing the outcome v on the 5 most marginally correlated q_{jL}, ie, the 5 which have the highest $|\widehat{\text{corr}}(v, q_{jL}(z)|$. Such modification was also used in [3].

A.1.2. *Penalty level choice for single outcome.* In every case when a single outcome variable is considered in isolation (this includes the reduced form selection step and the selection step corresponding to Φ_{K1}), Lasso is implemented with penalty λ as described in [1]. For ease of reference, note that [1] suggest λ given by $2c_{\text{Lasso}}F^{-1}_{\text{N}(0,1)}(1-\alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} > 1, \alpha_{\text{Lasso}} \to 0$ are tuning parameters. In every instance in this paper, $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05$ are used.

A.1.3. *Penalty level choice for $\Phi_{K,\text{Simple}}$.* In this case, K Lasso regressions are run simultaneously. In this case, for all $\varphi \in \Phi_K$, λ is given by $2c_{\text{Lasso}}F^{-1}_{\text{N}(0,1)}(1-\alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05/K$ are used.

A.1.4. *Penalty level choice and implementation for $\Phi_{K,\text{Span}}$.* When the Span option is used, $\Phi_{K,Span}$ is decomposed $\Phi_{K,\text{ Span}} = \Phi_{K1} \cup \Phi_{K2} \cup \Phi_{K3}$. Each component has a corresponding penalty level applied to all φ within that component. On the first component, $\lambda_{\Phi_{K_1}} = 2c_{\text{Lasso}}F^{-1}_{\text{N}(0,1)}(1-\alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05$. On the second component, $\lambda_{\Phi_{K_2}} = 2c_{\text{Lasso}}F^{-1}_{\text{N}(0,1)}(1-\alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05/K$. On the third component, $\lambda_{\Phi_{K_3}} = 2c_{\text{Lasso}}F^{-1}_{\text{N}(0,1)}(1-\alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05/K$.

The following procedure is used for approximating I_{Φ_K} in the case that a component of Φ_K contains a continuum of test functions. For each $j \leqslant L$, a Lasso regression $\check{\varphi}_j \in \Phi_{K3}$ which is more likely to select $q_{jL(z)}$ than other $\varphi \in \Phi_K$. Specifically, for each j, $\check{\varphi}_j$ is set to the linear combination of $p_{1K}, ..., p_{KK}$ with highest marginal correlation to q_{jL}. Then the approximation to the first stage model selection step proceeds by using $\check{I}_{\Phi_{K3}} = \bigcup_{j \leqslant L} I_{\check{\varphi}_j(x)}$ in place of $I_{\Phi_{K3}}$.

A.1.5. *Penalty level choice for* $\Phi_{K,\text{Span-Conservative}}$. When the Conservative Span option is used, $\Phi_{K,\text{ Span-Conservative}}$ is decomposed $\Phi_{K,\text{ Span-Conservative}} = \Phi_{K1} \cup \Phi_{K2} \cup \Phi_{K3}$. Each component again has a corresponding penalty level applied to all φ within that component. On the first component, $\lambda_{\Phi_{K1}} = 2c_{\text{Lasso}} F^{-1}_{\text{N}(0,1)}(1 - \alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05$. On the second component, $\lambda_{\Phi_{K2}} = 2c_{\text{Lasso}} F^{-1}_{\text{N}(0,1)}(1 - \alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01$ and $\alpha_{\text{Lasso}} = .05/K$. On the third component, $\lambda_{\Phi_{K3}} = 2c_{\text{Lasso}} F^{-1}_{\text{N}(0,1)}(1 - \alpha_{\text{Lasso}}/L)$ where $c_{\text{Lasso}} = 1.01K^{1/2}$ and $\alpha_{\text{Lasso}} = .05$.

In order to approximate the variables selected on the continuum of Lasso estimates indexed by Φ_{K3}, the identical procedure with the Span option above is used. Note that the only difference between the Conservative Span option and the Span option is in $\lambda_{\Phi_{K3}}$.

A.2. p^K **Implementation Details.** In every simulation, p^K is constructed using a cubic B-spline expansion. For fixed K, the approximating dictionary is chosen according to the following procedure. Knots points $t_1, ..., t_{K-3}$ are chosen according to the following rule. Set

$$t_{\max} = \text{quantile}_{0.95}(|x_1|, ..., |x_n|) \text{ and } t_{\min} = -t_{\max}.$$

Let $\Delta_k = t_k - t_{k-1}$. For constants $c_1, c_2 \geqslant 0$ set

$$\Delta_k = c_1 + c_2|(K-2)/2 - k|$$

for $k = 2, ..., K-3$.

The constants c_1, c_2 serve to insert more knot points where the density of x is higher. The choices for c_1, c_2 are determined uniquely by the condition that $c_1 = 2c_2$ and that the endpoints satisfy $t_1 = t_{\min}$ and $t_{K-3} = t_{\max}$. Next, the B-spline formulation used here is given by the recursive formulation. Set

$$B_{k,0}(x) = \mathbf{1}_{t_k \leqslant x < t_{k+1}}.$$

Set $B_{k,0} = 0$ for k outside of $1, ..., K-3$. In addition, for spline order $o > 0$,

$$B_{k,o}(x) = \frac{x - t_k}{t_{k+o} - t_k} B_{k,o-1} + \frac{t_{k+o+1} - x}{t_{k+o+1} - t_{k+1}} B_{k+1,o-1}.$$

Set $(p_{1,K}(x), ..., p_{K-3,K}(x)) = (B_{1,3}(x), ...B_{K-3,3}(x))$. The dictionary is completed by adding the additional terms $p_{K-2,K}(x) = x,\ p_{K-1,K}(x) = x^2,\ p_{K,K}(x) = x^3$.

$\widehat{K}$ is chosen according to the following procedure. First, an initial set of terms $q^{\text{initial}}(z) \subseteq q^L(z)$ is selected. In each case, $q^{\text{initial}}(z)$ contains the terms I_{RF}. That is, the terms selected in a Lasso regression y on $q^L(z)$. Next, an initial value $\widehat{K}_0 \leqslant 2\lfloor n^{1/3} \rfloor$ is chosen to minimize BIC using $(p^K(x), q^{\text{initial}}(z))$. In the simulation study, $\widehat{K}_0$ is constrained to be $\geqslant 5$. Finally, in order to ensure undersmoothing, $\widehat{K}$ is set to $\widehat{K} = \lfloor (\log_{10}(n))\widehat{K}_0 \rfloor$ in the simulation studies and $\widehat{K} = \widehat{K}_0 + 1$ in the empirical example.

In the empirical example, separate components of g_0, given by $g_0^{\text{math}}(\text{ACT}_{\text{math}}), g_0^{\text{eng}}(\text{ACT}_{\text{eng}}), g_0^{\text{read}}(\text{ACT}_{\text{read}} g_0^{\text{sci}}(\text{ACT}_{\text{sci}}))$ are each approximated with separate dictionaries $p_{\text{math}}^{K_{\text{math}}}$, $p_{\text{eng}}^{K_{\text{eng}}}$, $p_{\text{read}}^{K_{\text{read}}}$, $p_{\text{sci}}^{K_{\text{sci}}}$. The restriction $K_{\text{math}} = K_{\text{eng}} = K_{\text{read}} = K_{\text{sci}}$ is enforced. Set

$$K = K_{\text{math}} + K_{\text{eng}} + K_{\text{read}} + K_{\text{sci}}.$$

Each of the above four dictionaries is a B-spline basis defined exactly as in the simulations. $\widehat{K}$ is chosen to minimize BIC in the same way as in simulations. Then $\widehat{K}$ is set to

$$\widehat{K} = 4 + \widehat{K}_0.$$

The choice $\widehat{K} = 4 + \widehat{K}_0$ in the empirical example, instead of $\widehat{K} = \lfloor(\log_{10}(n))\rfloor$, is made to avoid $\widehat{K} > 4 \times 36$ which corresponds to the size of the support of the data.

A.3. **Targeted Undersmoothing Implementation Details.** The following procedure is used to estimate the Targeted Undersmoothing (TU; specifically TU(1); see [3]) confidence intervals for θ_0. For each $I \subseteq \{1, ..., p\}$ let $\widehat{\text{CI}}_{K,I}(\theta_0)$ be the corresponding confidence interval for θ_0 using K terms and the components of q^L corresponding to I. Then the full TU confidence interval is defined by the convex hull of $\cup_{j \leqslant p} \widehat{\text{CI}}_{\widehat{K}, I_{\text{RF}} \cup \{j\}}(\theta_0)$. In this implementation, a truncated TU confidence interval is calculated instead: $\cup_{j \leqslant 2s_0} \widehat{\text{CI}}_{\widehat{K}, I_{\text{RF}} \cup \{j\}}(\theta_0)$. This is done because the simulation run time reduces to the order of a day (from the order of a month), and therefore helps facilitate easier replicability. Changing the code to calculate the full TU confidence intervals is trivial. This also highlights that computing speed is another advantage of the Post-Nonparametric Double procedure relative to TU in certain settings. In terms of approximation error, the full TU estimator was implemented for the case $n = 100$, $p = 50$ for 1000 replications. The full TU confidence intervals as well as the truncated TU confidence intervals each made 9 false rejections. In addition, the average interval length for the full TU intervals was 1.740 while the average interval length for the truncated TU intervals was 1.722. Therefore, the truncated and full TU confidence intervals show very similar performance in this instance.

Appendix B. Proofs

B.1. **Preliminary Setup and Additional Notation.** Throughout the course of the proof, as much reference as possible is made to results in [4], [2]. This is done in order to maximize clarity and to present a better picture of the overall argument. In many cases, appealing directly to arguments in [4] is possible because many of the bounds required for deriving asymptotic normality for series estimators depend only on properties of $\widehat{g}$, g_0, p^K and D. Less direct appeal to bounds in the original Post-Double Selection argument is possible, since those arguments do not track K, and do not have notions of quantities stemming from Φ_K like α_ρ, α_Φ. However, the main idea of decomposing p^K into components in the span of, and orthogonal to q^L, remains as a theme throughout the proofs.

For any function φ, let $\varphi(X)$ denote the vector $[\varphi(x_1), \varphi(x_2), ..., \varphi(x_n)]'$. Similarly, let $\phi_{q^L}\varphi(Z) = [\pi_{q^L}\varphi(z_1), \pi_{q^L}\varphi(z_2), ..., \pi_{q^L}\varphi(z_n)]'$. In addition, define the following quantities.

1. Let m be the $n \times K$ matrix $m = \pi_{q^L} p^K(Z) = [\pi_{q^L} p_{1K}(Z), ..., \pi_{q^L} p_{KK}(Z)]$
2. Let $W = P - m$
3. Let $\widehat{\Omega} = n^{-1} P' \mathcal{M} P$
4. Let $\Omega = n^{-1} \mathrm{E}[W'W]$
5. Let $\bar{\Omega} = n^{-1} W'W$
6. Let m be partitioned $m = [m_1, ..., m_K]$
7. Let W be partitioned $m = [W_1, ..., W_K]$
8. For any $\varphi \in \Phi_K$, let $R_\varphi = Q(\beta_{\varphi,L} - \beta_{\varphi,L,s_0})$
9. Let $R_y = Q(\beta_{y,L} - \beta_{y,L,s_0})$
10. For any φ, let $U_\varphi = \varphi(X) - Q\beta_{\varphi,L}$
11. Let $U_y = Y - Q\beta_{y,L}$
12. Let $F = V^{-1/2}$
13. Let $\varphi_a(x)$ be the function such that $\pi_{q^L} \varphi_a(Z) = FA'm$
14. Let $m_a = FA'm$
15. Let $W_a = \varphi_a(X) - m_a$.
16. For $g \notin \Phi_K$, let $R_g = \pi_{q^L} g(Z) - \eta_1 Q(\beta_{\varphi_1,L} - \beta_{\varphi_1,L,s_0}) - ... - \eta_{k_g} Q(\beta_{\varphi_{k_g},L} - \beta_{\varphi_{k_g},L,s_0})$ for some $(\varphi_1, ...\varphi_{k_g}), (\eta_1, ..., \eta_{k_g})$ within a fixed constant factor of achieving the infinum in the density assumption (Assumption 8.)
17. Let $R_m = [R_{m_1}, ..., R_{m_K}]$.

Assume without loss of generality that $B_K = \mathrm{Id}_K$, the identity matrix of order K. The reason this is without loss of generality is that dictionary p^K is used only in the post-selection estimation, while Φ_K is used for first stage model selection. In addition, assume without loss of generality that $\Omega = \mathrm{Id}_K$.

Throughout the exposition, there is a common naming convention for various regression coefficients. Quantities of the form $\widehat{\beta}_{v,I}$ always denotes the sample regression coefficients from regressing the variable v on the components specified by I. This implies that the quantities $\widehat{\beta}_{\varphi,I_{\varphi,L}} = \widehat{\beta}_{\varphi,L,\text{Post-Lasso}}$ are equivalent, since the specified components being regressed on are the same. In addition, $\widehat{\beta}_{\varphi,I_{\Phi_K+\mathrm{RF}}} = \widehat{\beta}_{\varphi,\tilde{q}} = \widehat{\beta}_{\varphi(X),I_{\Phi_K+\mathrm{RF}}}$ are equivalent. Next, quantities of the form $\beta_{v,L}$ and β_{v,L,s_0} without a hat accent are population quantities and are defined in the text above.

B.2. **Preliminary Lemmas.**

Lemma 1. *Under the assumptions of Theorem 1,*

1. $J_1 := \max_{k \leqslant K} n^{-1/2} \|Q'W_k\|_\infty = O_p(\big(\log(KL)^{1/2}\big))$
2. $J_2 := n^{-1/2} \|Q'\mathcal{E}\|_\infty = O_p(\big(\log(L)^{1/2}\big))$
3. $J_3 := n^{-1/2} \|R_m'\mathcal{E}\|_2 = O_p((KK^{\alpha_\rho} L^{-\alpha_z}))$
4. $J_4 := n^{-1/2} \|R_{h_0}'W\|_2 = O_p(\big(n^{1/2} \zeta_0(K) L^{-\alpha_z}\big))$

5. $J_5 := \max_{k \leqslant K} n^{-1/2} \|\mathcal{M} m_k\|_2$
$$= O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0^{1/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$$
6. $J_6 := n^{-1/2} \|\mathcal{M} h_0(Z)\|_2 = O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0^{1/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$
7. $J_7 := \max_{k \leqslant K} \|\widehat{\beta}_{m_k, I_{\Phi_K + RF}} - \beta_{p_{kK}, L, s_0}\|_1$
$$= O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0 K^{\alpha_{I_\Phi}/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$$
8. $J_8 := \|\widehat{\beta}_{h_0, I_{\Phi_K + RF}} - \beta_{h_0, L, s_0}\|_1$
$$= O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0 K^{\alpha_{I_\Phi}/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$$
9. $J_9 := \max_{k \leqslant K} \|\widehat{\beta}_{W_k, I_{\Phi_K + RF}}\|_1 = O_p(\Big(n^{-1/2} s_0^{1/2} K^{\alpha_{I_\Phi}/2} \log(KL)^{1/2}\Big))$
10. $J_{10} := \|\widehat{\beta}_{\mathcal{E}, I_{\Phi_K + RF}}\|_1 = O_p(\Big(n^{-1/2} s_0^{1/2} K^{\alpha_{I_\Phi}/2} \log(L)^{1/2}\Big))$
11. $J_{11} := n^{-1/2} \|Q' W_a\|_\infty = O_p(\big(\log(KL)^{1/2}\big))$
12. $J_{12} := n^{-1/2} \|R'_{m_a} \mathcal{E}\|_2 = O_p((K K^{\alpha_\rho} L^{-\alpha_z}))$
13. $J_{13} := n^{-1/2} \|\mathcal{M} m_a\|_2 = O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0^{1/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$
14. $J_{14} := \|\widehat{\beta}_{m_a, I_{\Phi_K + RF}} - \beta_{\varphi_a, L, s_0}\|_1$
$$= O_p(\Big(n^{-1/2} K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0 K^{\alpha_{I_\Phi}/2} \log(L)^{1/2} + L^{-\alpha_z} K^{\alpha_\rho}\Big))$$
15. $J_{15} := \|\widehat{\beta}_{W_a, I_{\Phi_K + RF}}\|_1 = O_p(\Big(n^{-1/2} s_0^{1/2} K^{\alpha_{I_\Phi}/2} \log(KL)^{1/2}\Big))$
16. $J_{16} := n^{-1} \|R'_m W\|_{\mathcal{F}} = O_p(K^{1/2} \zeta_0(K) K^{\alpha_\rho} L^{-\alpha_z})$.

Proof.

Statement 1. By Lemma 5 of [2], two conditions which together are sufficient for $\max_{k \leqslant K j \leqslant L} \frac{|Q'_j W_k|}{\sqrt{\sum_{i=1}^n q_{jL}(z_i)^2 W_{ki}^2}} = O_p((\log KL)^{1/2})$ are that $\max_{k \leqslant K, j \leqslant L} \frac{\mathrm{E}[|q_{jL}(z)|^3 |W_{ik}|^3]^{1/3}}{\mathrm{E}[q_{jL}(z)^2 W_{ik}^2]^{1/2}} = O(\zeta_0(K))$ and the rate condition $\log KL = o(\zeta_0(K)^{-1} n^{1/3})$. Note that $\mathrm{E}[q_{jL}(z)^2 W_{ik}^2]^{1/2}$ is bounded away from zero by assumption. In addition, by Hölder's inequality, $\mathrm{E}[|q_{jL}(z)|^3 |W_{ik}|^3] \leqslant \mathrm{E}[|q_{jL}(z)|^3] \zeta_0(K)^3$. This implies that the first condition holds. The second condition is given in the assumptions.

Statement 2. Follows similarly as Statement 1.

Statement 3. This statement follows directly from the fact that $\mathrm{E}[\varepsilon|x,z]=0$, $\mathrm{E}[\varepsilon^2|x,z]$ bounded, along with $\dim(R_m'\mathcal{E})=K$ and $\|R_{m_k}\|_\infty = O(K^{\alpha_\rho}L^{-\alpha_z})$ by the density assumption, allowing the use of the K-dimensional Chebyshev Inequality.

Statement 4. $\|R_{h_0}'W\|_2 = \|\sum_i R_{h_0,i}W_i\|_2 \leqslant O(L^{-\alpha_z})\zeta_0(K)$ by the facts that $\|R_{h_0}\|_\infty = O(L^{-\alpha_z})$ and $\|W_i\|_2 \leqslant \zeta_0(K)$.

Statement 5.

First note that the following two hold.

1. For any $\varphi \in \Phi_K$, $\mathcal{M}\pi_{q^L}\varphi(Z) = \mathcal{M}R_\varphi + \mathcal{M}(Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\varphi,L}})$.
2. For any $g \in \mathrm{LinSpan}(p^K)$, and any corresponding expansion $g = \eta_1\varphi_1 + ... + \eta_{k_g}\varphi_{k_g} + r_g$ with $\eta_1,...\eta_{k_g} \in \mathbb{R}, \varphi_1,...\varphi_{k_g} \in \Phi_K$,

$$\|\mathcal{M}\pi_{q^L}g(Z)\|_2 \leqslant \|\eta\|_1 \max_{\varphi\in\{\varphi_1,...,\varphi_{k_g}\}} (\|Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\varphi,L}}\|_2 + \|R_\varphi\|_2) + \|r_g(Z)\|_2.$$

To show the first of the above two statements, for each $\varphi \in \Phi_K$, note that

$$\begin{aligned}
\mathcal{M}\pi_{q^L}\varphi(Z) &= \mathcal{M}\mathcal{M}\pi_{q^L}\varphi(Z)\\
&= \mathcal{M}(\pi_{q^L}\varphi(Z) - \mathcal{P}\pi_{q^L}\varphi(Z))\\
&= \mathcal{M}(Q\beta_{\varphi,L} - \mathcal{P}(\varphi(X) - U_\varphi))\\
&= \mathcal{M}(Q\beta_{\varphi,L} - Q\widehat{\beta}_{\varphi,I_{\Phi_K+\mathrm{RF}}} + \mathcal{P}U_\varphi)\\
&= \mathcal{M}R_\varphi + \mathcal{M}(Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\Phi_K+\mathrm{RF}}}) + \mathcal{M}\mathcal{P}U_\varphi\\
&= \mathcal{M}R_\varphi + \mathcal{M}(Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\Phi_K+\mathrm{RF}}})\\
&= \mathcal{M}R_\varphi + \mathcal{M}(Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\varphi,L}}) + \underbrace{\mathcal{M}(Q\widehat{\beta}_{\varphi,I_{\varphi,L}} - Q\widehat{\beta}_{\varphi,I_{\Phi_K+\mathrm{RF}}})}\\
&\qquad\qquad = \mathcal{M}(\mathcal{P}_{I_{\varphi,L}}\varphi(X) - \mathcal{P}\varphi(X))\\
&\qquad\qquad = \mathcal{M}\mathcal{P}(\mathcal{P}_{I_{\varphi,L}}\varphi(X) - \varphi(X))\\
&\qquad\qquad = 0
\end{aligned}$$

$$\Rightarrow \mathcal{M}\pi_{q^L}\varphi(Z) = \mathcal{M}R_\varphi + \mathcal{M}(Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\varphi,L}}).$$

This establishes the first claim. Now turn to the second claim. Note that using the density assumption, there are $\varphi_1,...,\varphi_{k_g}$ and a vector $\eta = (\eta_1,...,\eta_{k_g})$ such that $g = \eta_1\varphi_1 + ... + \eta_{k_g}\varphi_{k_g} + r_g$ for some remainder r_g, sufficiently small. Then

$$\|\mathcal{M}\pi_{q^L}g(Z)\|_2 = \|\eta_1\mathcal{M}\pi_{q^L}\varphi_1(X) + ... + \eta_{k_g}\mathcal{M}\pi_{q^L}\varphi_{k_g}(Z) + \mathcal{M}\pi_{q^L}r_g(Z)\|_2$$

Next, looking at each φ in the above expansion (ie each $\varphi \in \{\varphi_1,...,\varphi_{k_g}\}$) and combining the above expression gives

$$\|\mathcal{M}\pi_{q^L} g(Z)\|_2 = \|\eta_1 \mathcal{M} R_{\varphi_1} + \eta_1 \mathcal{M}(Q\beta_{\varphi_1,L,s_0} - Q\widehat{\beta}_{\varphi_1,I_{\varphi_1}}) + ...$$
$$... + \eta_{k_g} \mathcal{M} R_{\varphi_{k_g}} + \eta_{k_g} \mathcal{M}(Q\beta_{\varphi_{k_g},L,s_0} - Q\widehat{\beta}_{\varphi_{k_g},I_{\varphi_{k_g}}}) + \mathcal{M}\pi_{q^L} r_g(Z)\|_2.$$

Applying Hölder's inequality and the fact that $\mathcal{M}$ is a projection (and hence non-expansive) gives the bound

$$\leqslant \|\eta\|_1 \max_{\varphi \in \{\varphi_1,\dots,\varphi_{k_g}\}} (\|Q\beta_{\varphi,L,s_0} - Q\widehat{\beta}_{\varphi,I_{\varphi,L}}\|_2 + \|R_\varphi\|_2) + \|r_g(Z)\|_2.$$

These can then be applied directly to $n^{-1/2}\|\mathcal{M}m_k\|_2$. The corresponding η and R_{m_k} satisfy $L^{-\alpha_Z}\|\eta\|_1 \leqslant O(K^{-\alpha_\rho})$ and $\|R_{m_k}\|_2 \leqslant n^{1/2}O(K^{-\alpha_\rho})$. Then we have the bound

$$\|\mathcal{M}\pi_{q^L} g(Z)\|_2 = O_p(K^{\alpha_\rho} K^{\alpha_\Phi/2} s_0^{1/2} \log(L)^{1/2} + n^{1/2} K^{-\alpha_\rho}).$$

Under Assumption 10, note that for each m_k, taking $\eta = 1$ and $R_{m_k} = 0$ are feasible by assumption. The result follows.

Statement 6.

$$\begin{aligned}
n^{-1/2}\|\mathcal{M}h_0(Z)\|_2 &= n^{-1/2}\|\mathcal{M}(Q\beta_{h_0,L,s_0} + R_{h_0})\|_2) \\
&\leqslant n^{-1/2}(\|\mathcal{M}Q\beta_{h_0,L,s_0}\|_2 + \|\mathcal{M}R_{h_0}\|_2) \\
&\leqslant n^{-1/2}(\|\mathcal{M}Q(\beta_{g_0,L,s_0} + \beta_{h_0,L,s_0} - Q\beta_{g_0,L,s_0})\|_2 + \|\mathcal{M}R_{h_0})\|_2 \\
&= n^{-1/2}(\|\mathcal{M}(Q\beta_{y,L,s_0} - Q\beta_{g_0,L,s_0})\|_2 + \|\mathcal{M}R_{h_0}\|_2) \\
&\leqslant n^{-1/2}(\|\mathcal{M}Q\beta_{y,L,s_0}\|_2 + \|\mathcal{M}Q\beta_{g_0,L,s_0}\|_2 + \|\mathcal{M}R_{h_0}\|_2) \\
&= n^{-1/2}(\|\mathcal{M}\pi_{q^L} y(Z)\|_2 + \|\mathcal{M}\pi_{q^L} g_0(Z)\|_2 + \|\mathcal{M}R_{h_0}\|_2)
\end{aligned}$$

The first two terms above, $n^{-1/2}(\|\mathcal{M}Q\pi_{q^L} y(Z)\|_2 + n^{-1/2}\|\mathcal{M}\pi_{q^L} g_0(Z)\|_2)$, are $O_p(K^{\alpha_\rho} K^{\alpha_\Phi/2} n^{-1/2} s_0^{1/2} \log(L)^{1/2} + L^{-\alpha_Z} K^{\alpha_\rho})$ by the same reasoning as Statement 5. In addition $n^{-1/2}\|\mathcal{M}R_{h_0}\|_2 \leqslant n^{-1/2}\|R_{h_0}\|_2 = O(L^{-\alpha_Z})$ by assumption. This gives

$$n^{-1/2}\|\mathcal{M}h_0(Z)\|_2 = O_p(K^{\alpha_\rho} K^{\alpha_\Phi/2} n^{-1/2} s_0^{1/2} \log(L)^{1/2} + L^{-\alpha_Z} K^{\alpha_\rho}).$$

Statement 7.

$$\begin{aligned}
&\|\widehat{\beta}_{m_k, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0}\|_1 \\
&\leqslant |I_{\Phi_K+\mathrm{RF}}|^{1/2}\|\widehat{\beta}_{m_k, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0}\|_2 \\
&= |I_{\Phi_K+\mathrm{RF}}|^{1/2}\left((\widehat{\beta}_{m_k, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0})'(\widehat{\beta}_{m_k, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0})\right)^{1/2} \\
&\leqslant |I_{\Phi_K+\mathrm{RF}}|^{1/2}O_p(1)\left((\widehat{\beta}_{m_k, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0})'(Q'_{I_{\Phi_K+\mathrm{RF}}}Q_{I_{\Phi_K+\mathrm{RF}}}/n)(\widehat{\beta}_{g, I_{\Phi_K+\mathrm{RF}}} - \beta_{p_{kK},L,s_0})\right)^{1/2} \\
&= |I_{\Phi_K+\mathrm{RF}}|^{1/2}O_p(1)n^{-1/2}\|\mathcal{P}m_k - Q\beta_{p_{kK},L,s_0}\|_2 \\
&= |I_{\Phi_K+\mathrm{RF}}|^{1/2}O_p(1)n^{-1/2}\|m_k - \mathcal{M}m_k - Q\beta_{p_{kK},L,s_0}\|_2 \\
&= |I_{\Phi_K+\mathrm{RF}}|^{1/2}O_p(1)n^{-1/2}\| - \mathcal{M}m_k + R_{m_k}\|_2 \\
&\leqslant |I_{\Phi_K+\mathrm{RF}}|^{1/2}O_p(1)(J_5 + O(L^{-\alpha_z})) \\
&= O_p(s_0^{1/2}K^{\alpha_{I_\Phi}/2})(J_5 + O(L^{-\alpha_z})) \\
&= O_p(s_0^{\alpha_{I_\Phi}/2+1/2})O_p(K^{\alpha_\rho}K^{\alpha_\Phi/2}s_0^{1/2}\log(L)^{1/2} + n^{1/2}L^{-\alpha_z}K^{\alpha_\rho}) \\
&= O_p(K^{\alpha_\rho}K^{\alpha_\Phi/2}s_0^1K^{\alpha_{I_\Phi}/2}\log(L)^{1/2} + n^{1/2}L^{-\alpha_z}K^{\alpha_\rho}).
\end{aligned}$$

Statement 8. Proven analogously to Statement 7.

Statement 9.

$$\begin{aligned}
&\max_{k\leqslant K}\|\widehat{\beta}_{W_k, I_{\Phi_K+\mathrm{RF}}}\|_1 \\
&= \max_{k\leqslant K}\|(\tilde{Q}'\tilde{Q})^{-1}\tilde{Q}'W_k\|_1 \\
&\leqslant ||I_{\Phi_K+\mathrm{RF}}|^{1/2}\max_{k\leqslant K}\|(\tilde{Q}'\tilde{Q})^{-1}\tilde{Q}'W_k\|_2 \\
&\leqslant |I_{\Phi_K+\mathrm{RF}}|^{1/2}\kappa_{\min}^{-1/2}(I_{\Phi_K+\mathrm{RF}}|)\max_{k\leqslant K}\|n^{-1}Q'W_k\|_\infty \\
&= O_p(s_0^{1/2}K^{\alpha_{I_\Phi}/2}\cdot 1\cdot n^{-1/2}\log(KL)^{1/2}).
\end{aligned}$$

Statement 10. Proven analogously to Statement 9.

Statements 11-15. Proven analogously to Statements 1,3,5,7,9.

Statement 16.

$$n^{-1}\Big\|\sum_{i=1}^n W_i'R_{m,i}\Big\|_{\mathcal{F}} = n^{-1}\left(\sum_k \|W'R_{m_k}\|_2^2\right)^{1/2}$$

$$\leqslant n^{-1}\left(\sum_k n^2\zeta_0(K)^2\|R^2_{m_k}\|_\infty\right)^{1/2}.$$

By the density assumption, $\|R_{m_k}\|_\infty \leqslant K^{\alpha_\rho}L^{-\alpha_z}$. This then implies that

$$n^{-1}\Big\|\sum_{i=1}^n W_i' R_{m,i}\Big\|_{\mathcal{F}} \leqslant K^{1/2}K^{-\alpha_\rho}.$$

□

Lemma 2.

1. $\Xi_1 := n^{-1}\|W'\mathcal{P}W\|_{\mathcal{F}} \leqslant n^{-1/2}KJ_9J_1$
2. $\Xi_2 := n^{-1}\|m'\mathcal{M}m\|_{\mathcal{F}} \leqslant KJ_5^2$
3. $\Xi_3 := n^{-1}\|m'\mathcal{M}W\|_{\mathcal{F}} \leqslant J_{16} + n^{-1/2}KJ_7J_1$
4. $\Xi_4 := n^{-1/2}\|m'\mathcal{M}h_0(Z)\|_2 \leqslant n^{1/2}K^{1/2}J_5J_6$
5. $\Xi_5 := n^{-1/2}\|W'\mathcal{M}h_0(Z)\|_2 \leqslant J_4 + K^{1/2}J_8J_1$
6. $\Xi_6 := n^{-1/2}\|W'\mathcal{P}\mathcal{E}\|_2 \leqslant K^{1/2}J_9J_2$
7. $\Xi_7 := n^{-1/2}\|m'\mathcal{M}\mathcal{E}\|_2 \leqslant J_4 + K^{1/2}J_7J_2$
8. $\Xi_8 := n^{-1/2}|m_a'\mathcal{M}h_0(Z)| \leqslant n^{1/2}J_5J_{13}$
9. $\Xi_9 := n^{-1/2}|W_a'\mathcal{M}h_0(Z)| \leqslant J_{12} + J_{14}J_1$
10. $\Xi_{10} := n^{-1/2}|W_a'\mathcal{P}\mathcal{E}| \leqslant J_9J_{11}$
11. $\Xi_{11} := n^{-1/2}|m_a'\mathcal{M}\mathcal{E}| \leqslant J_{12} + J_7J_{11}$.

Proof.

Statement 1.

$$\begin{aligned}
\left(n^{-1}\|W'\mathcal{P}W\|_{\mathcal{F}}\right)^2 &= \sum_{k,\bar{k}\leqslant K}(n^{-1}W_k'\mathcal{P}W_{\bar{k}})^2 = \\
&= \sum_{k,\bar{k}\leqslant K}(n^{-1}\widehat{\beta}'_{W_k,I_{\Phi_K+RF}}Q'W_{\bar{k}})^2 \\
&\leqslant \sum_{k,\bar{k}\leqslant K}\|n^{-1/2}\widehat{\beta}_{W_k,I_{\Phi_K+RF}}\|_1^2\|n^{-1/2}Q'W_{\bar{k}}\|_\infty^2 \\
&= \left(\sum_{k\leqslant K}\|n^{-1/2}\widehat{\beta}_{W_k,I_{\Phi_K+RF}}\|_1^2\right)\left(\sum_{\bar{k}\leqslant K}\|n^{-1/2}Q'W_{\bar{k}}\|_\infty^2\right) \\
&\leqslant K\cdot n^{-1}J_9^2\cdot K\cdot J_1^2 \\
\\
&\Rightarrow n^{-1}\|W'\mathcal{P}W\|_{\mathcal{F}} \leqslant n^{-1/2}KJ_1J_9.
\end{aligned}$$

Statement 2.

$$
\begin{aligned}
\left(n^{-1}\|m'\mathcal{M}m\|_{\mathcal{F}}\right)^2 &= \sum_{k,\bar{k}\leqslant K}(n^{-1}m_k'\mathcal{M}m_{\bar{k}})^2 \leqslant \sum_{k,\bar{k}\leqslant K}\|n^{-1/2}\mathcal{M}m_k\|_2^2\|n^{-1/2}\mathcal{M}m_{\bar{k}}\|_2^2 \\
&= \left(\sum_{k\leqslant K}\|n^{-1/2}\mathcal{M}m_k\|_2^2\right)^2 \leqslant K^2J_5^4 \\
&\Rightarrow n^{-1}\|m'\mathcal{M}m\|_{\mathcal{F}} \leqslant KJ_5^2.
\end{aligned}
$$

Statement 3.

$$
\begin{aligned}
n^{-1}\|m'\mathcal{M}W\|_{\mathcal{F}} &= n^{-1}\|m'W/n - m'\mathcal{P}W\|_{\mathcal{F}} \\
&= n^{-1}\|R_m'W + (Q\beta_{p^K,L,s_0})'W - m'\mathcal{P}W\|_{\mathcal{F}} \\
&= n^{-1}\|R_m'W + (Q\beta_{p^K,L,s_0})'W - (Q\widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'W\|_{\mathcal{F}} \\
&= n^{-1}\|R_m'W + (\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'W\|_{\mathcal{F}} \\
&\leqslant n^{-1}\|R_m'W\|_{\mathcal{F}} + n^{-1}\|(\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'W\|_{\mathcal{F}}.
\end{aligned}
$$

Then the first term in the last line is bounded above as $n^{-1}\|R_m'W\|_{\mathcal{F}} = J_{16}$ while the second term has

$$
\begin{aligned}
&\left(n^{-1}\|(\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'W\|_{\mathcal{F}}\right)^2 \\
&= n^{-2}\sum_{k,\bar{k}\leqslant K}((\beta_{p_k,L,s_0} - \widehat{\beta}_{m_k,I_{\Phi_K+\mathrm{RF}}})'Q'W_{\bar{k}})^2 \\
&\leqslant n^{-2}\sum_{k,\bar{k}\leqslant K}\|\beta_{p_k,L,s_0} - \widehat{\beta}_{m_k,I_{\Phi_K+\mathrm{RF}}}\|_1^2\|Q'W_{\bar{k}}\|_\infty^2 \\
&= n^{-1}\left(\sum_{k\leqslant K}\|\beta_{p_k,L,s_0} - \widehat{\beta}_{m_k,I_{\Phi_K+\mathrm{RF}}}\|_1^2\right)\left(\sum_{\bar{k}\leqslant K}\|n^{-1/2}Q'W_{\bar{k}}\|_\infty^2\right) \\
&\leqslant n^{-1}K\cdot J_7^2\cdot K\cdot J_1^2.
\end{aligned}
$$

Therefore, $n^{-1}\|m'\mathcal{M}W\|_{\mathcal{F}} \leqslant J_{16} + n^{-1/2}KJ_7J_1$.

Statement 4.

$$
\begin{aligned}
n^{-1/2}\|m'\mathcal{M}h_0(Z)\|_2 &\leqslant n^{1/2}\|n^{-1/2}\mathcal{M}h_0(Z)\|_2K^{1/2}\max_{k\leqslant K}n^{-1/2}\|m'\mathcal{M}\|_2 \\
&\leqslant n^{1/2}K^{1/2}J_5J_6.
\end{aligned}
$$

Statement 5.

$$\begin{aligned}
n^{-1/2}\|W'\mathcal{M}h_0(Z)\|_2 &= n^{-1/2}\|W'h_0(Z) - W'\mathcal{P}h_0(Z)\|_2 \\
&= n^{-1/2}\|W'h_0(Z) - W'Q\widehat{\beta}_{h_0(Z),I_{\Phi_K+\mathrm{RF}}}\|_2 \\
&= n^{-1/2}\|W'R_{h_0} + W'Q\beta_{h_0,L,s_0} - W'Q\widehat{\beta}_{h_0(Z),I_{\Phi_K+\mathrm{RF}}}\|_2 \\
&= n^{-1/2}\left(\|W'h_0(Z)\|_2 + \|(\widehat{\beta}_{h_0(Z),I_{\Phi_K+\mathrm{RF}}} - \beta)'Q'W\|_2\right) \\
&\leqslant J_4 + K^{1/2}\max_{k\leqslant K}\|\widehat{\beta}_{h_0(Z),I_{\Phi_K+\mathrm{RF}}} - \beta_{h_0,L,s_0}\|_1\|n^{-1/2}Q'W_k\|_\infty \\
&\leqslant J_4 + K^{1/2}J_8J_1.
\end{aligned}$$

Statement 6.

$$\begin{aligned}
\left(n^{-1/2}\|W'\mathcal{P}W\|_2\right)^2 &= n^{-1}\sum_{k\leqslant K}(W_k'\mathcal{P}\mathcal{E})^2 \\
&= n^{-1}\sum_{k\leqslant K}(\widehat{\beta}'_{W_k,I_{\Phi_K+RF}}Q'\mathcal{E})^2 \\
&\leqslant \sum_{k\leqslant K}\|\widehat{\beta}_{W_k,I_{\Phi_K+RF}}\|_1^2\|n^{-1/2}Q'\mathcal{E}\|_\infty^2 \\
&\leqslant K\cdot J_9^2\cdot J_2^2 \\
&\Rightarrow n^{-1/2}\|W'\mathcal{P}\mathcal{E}\|_{\mathcal{F}} \leqslant K^{1/2}J_9J_2.
\end{aligned}$$

Statement 7.

$$\begin{aligned}
n^{-1/2}\|m'\mathcal{M}\mathcal{E}\|_2 &= n^{-1/2}\|m'\mathcal{E}/n - m'\mathcal{P}\mathcal{E}\|_{\mathcal{F}} \\
&= n^{-1/2}\|R_m'\mathcal{E} + (Q\beta_{p^K,L,s_0})'W - m'\mathcal{P}\mathcal{E}\|_{\mathcal{F}} \\
&= n^{-1/2}\|R_m'\mathcal{E} + (Q\beta_{p^K,L,s_0})'W - (Q\widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'\mathcal{E}\|_2 \\
&= n^{-1/2}\|R_m'\mathcal{E} + (\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'\mathcal{E}\|_2 \\
&\leqslant n^{-1/2}\|R_m'\mathcal{E}\|_2 + n^{-1/2}\|(\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'\mathcal{E}\|_2.
\end{aligned}$$

Then the first term in the last line is bounded above as $n^{-1/2}\|R_m'\mathcal{E}\|_2 = J_4$. Turning to the second term,

$$\begin{aligned}
&\left(n^{-1/2}\|(\beta_{p^K,L,s_0} - \widehat{\beta}_{m,I_{\Phi_K+\mathrm{RF}}})'Q'\mathcal{E}\|_2\right)^2 \\
&= n^{-1}\sum_{k\leqslant K}((\beta_{p_k,L,s_0} - \widehat{\beta}_{m_k,I_{\Phi_K+\mathrm{RF}}})'Q'\mathcal{E})^2 \\
&\leqslant \sum_{k\leqslant K}\|\beta_{p_k,L,s_0} - \widehat{\beta}_{m_k,I_{\Phi_K+\mathrm{RF}}}\|_1^2\|n^{-1/2}Q'\mathcal{E}\|_\infty^2 \\
&\leqslant K\cdot J_7^2\cdot J_2^2.
\end{aligned}$$

Therefore, $n^{-1/2}\|m'\mathcal{M}\mathcal{E}\|_2 \leqslant J_4 + K^{1/2}J_7J_2$.

Statements 8-11.

The argument is identical to the argument for Statements 4-7, adjusting appropriately for the fact that m_a is 1-dimensional rather than K-dimensional. □

The following corollaries follow directly from assumed rate conditions and the above bounds. These are used in the proof of Theorems 1 and 2.

Corollary 1. *Under the assumptions of Theorem 1,*

1. $\Xi_1 + \Xi_2 + \Xi_3 = O_p(n^{-1/2}\zeta_0(K)K^{1/2})$
2. $n^{-1/2}(\Xi_4 + \Xi_5 + \Xi_6 + \Xi_7) = O_p(n^{-1/2}K^{1/2} + K^{-\alpha_{g_0}})$.

Corollary 2. *Under the assumptions of Theorem 2,*

1. $n^{-1/2}\zeta_0(K)K^{1/2}(\Xi_4 + \Xi_5 + \Xi_6 + \Xi_7) = o_p(1)$.
2. $\Xi_8 + \Xi_9 + \Xi_{10} + \Xi_{11} = o_p(1)$.

Proof. Corollaries 1 and 2 follow from the rate conditions stated in Assumptions 2,9,11. The also follow from the following more general rate conditions.

Sufficient rate conditions for Corollary 1:

1. $s_0K^{\alpha_{I_\Phi}} = o(s_\kappa)$
2. $\log(KL) = o(\zeta_0(K)^{-1}n^{1/3})$
3. $L^{-\alpha_Z}n^{1/2}K^{-1/2}\zeta_0(K) = O(1)$
4. $L^{-2\alpha_Z}K^{2\alpha_\rho}(K^{1/2}n^{1/2}\zeta_0(K)^{-1} + n^{1/2} + K\log(L)^{1/2}\zeta_0(K)^{-2}) = O(1)$
5. $n^{-1/2}K^{1/2}s_0\log(L)\zeta_0(K)^{-1}(K^{2\alpha_\rho+\alpha_\Phi} + K^{\alpha_\rho+\alpha_\Phi/2+\alpha_{I_\Phi}/2}) = O(1)$
6. $n^{-1/2}s_0^{1/2}\log(L)(K^{2\alpha_\rho+\alpha_\Phi}s_0^{1/2} + K^{\alpha_{I_\Phi}/2}) = O(1)$.

Additional rate conditions sufficient (along with the above conditions) for Corollary 2:

1. $L^{-2\alpha_Z}K^{2\alpha_\rho}(\zeta_0(K)K + \zeta_0(K)^4K^{1-2\alpha_\rho} + K\log(L) + n^{1/2}) = o(1)$
2. $n^{-1}s_0\zeta_0(K)\log(L)(K^{1+2\alpha_\rho+\alpha_\Phi} + K^{1+\alpha_\rho+\alpha_\Phi/2+\alpha_{I_\Phi}/2}) = o(1)$
3. $n^{-1}s_0^2\log(L)^2(K^{4\alpha_\rho+2\alpha_\Phi} + K^{2\alpha_\rho+\alpha_\Phi+\alpha_{I_\Phi}}) = o(1)$
4. $s_0K^{\alpha_{I_\Phi}}\left(n^{-1/2}\zeta_0(K)K^{1/2} + K^{-\alpha_{g_0}}\right) = o(1)$
5. $n^{2/(4+\delta)}\zeta_0(K)n^{-1/2}K^{1/2} = o(1)$.

□

B.3. **Proof of Theorem 1.**

Lemma 3.

1. $\|\widehat{\Omega}-\Omega\|_{\mathcal{F}} \leqslant O_p(\zeta_0(K)K^{1/2}n^{-1/2}) + \Xi_1 + \Xi_2 + \Xi_3 = o_p(1)$
2. $\|\widehat{\Omega}^{-1}-\Omega^{-1}\|_{2\to 2} = O_p(\zeta_0(K)K^{1/2}n^{-1/2}) + O(\Xi_1 + \Xi_2 + \Xi_3)$.

Proof. The argument in Theorem 1 of [4] gives the bound $\|\bar{\Omega}-\Omega\|_{\mathcal{F}} = O_{\mathrm{P}}(\zeta_0(K)K^{1/2}n^{-1/2})$. Next, using the decomposition, $P = m + W$, write $\widehat{\Omega} = (m+W)'\mathcal{M}(m+W)/n = W'W/n - W'(\mathrm{Id}_n - \mathcal{M})W/n + m'\mathcal{M}m/n + 2m'\mathcal{M}W/n$. By triangle inquality, $\|\bar{\Omega}-\widehat{\Omega}\|_{\mathcal{F}} \leqslant \|W'\mathcal{P}W/n\|_{\mathcal{F}} + \|m'\mathcal{M}m/n\|_{\mathcal{F}} + \|2m'\mathcal{M}W/n\|_{\mathcal{F}} = \Xi_1+\Xi_2+\Xi_3$. Bounds for each of the three above terms are established above along with the assumed rate conditions give $\|\bar{\Omega}-\widehat{\Omega}\|_{\mathcal{F}} = o_p(1)$. The last statement holds by applying an expansion of the matrix inversion function around Id_K.

$$\widehat{\Omega}^{-1} = (\mathrm{Id}_K - (\mathrm{Id}_K - \widehat{\Omega}))^{-1} = \mathrm{Id}_K + (\mathrm{Id}_K - \widehat{\Omega}) + (\mathrm{Id}_K - \widehat{\Omega})^2 + ...$$

The sum given above is with probability $\to 1$ absolutely convergent relative to the Frobenius norm $\mathcal{F}$. In addition, by the bound $\|\cdot\|_{2\to 2} \leqslant \|\cdot\|_{\mathcal{F}}$, we have $\|\widehat{\Omega}^{-1} - \mathrm{Id}_K\|_{2\to 2} \leqslant \|\widehat{\Omega} - \mathrm{Id}_K\|_{\mathcal{F}} \leqslant \|\mathrm{Id}_K - \widehat{\Omega}\|_{\mathcal{F}} + \|\mathrm{Id}_K - \widehat{\Omega}\|^2_{\mathcal{F}} + ... = O_p(\zeta_0(K)K^{1/2}n^{-1/2}) + O(\Xi_1+\Xi_2+\Xi_3)$.

□

Note that since Ω has minimal eigenvalues bounded from below by assumption, it follows that $\widehat{\Omega}$ and $\bar{\Omega}$ are invertible with probability approaching 1. The reference [4] works on the event $1_n := \{\lambda_{\min}(\widehat{\Omega}) > 1/2\}$ and later uses the fact that this event has probability $\to 1$. This fact is used several times, however its use is only implicitly in reference to arguments in [4].

Lemma 4. $\|\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}\mathcal{E}\|_2 = O_p(n^{-1/2}K^{1/2})$.

Proof.

$$\begin{aligned}\|\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}\mathcal{E}\|_2 &\leqslant \|\widehat{\Omega}^{-1}\|_{2\to 2}n^{-1}\|P'\mathcal{M}\mathcal{E}\|_2 \\ &\leqslant \|\widehat{\Omega}^{-1}\|_{2\to 2}(n^{-1}\|W'\mathcal{E}\|_2 + n^{-1/2}\Xi_6 + n^{-1/2}\Xi_7)\end{aligned}$$

$\|W'\mathcal{E}\|_2 = O_p(n^{-1/2}K^{1/2} = O_p(n^{-1/2})$ by arguments in [4]. Bounds for $n^{-1/2}\Xi_6 + n^{-1/2}\Xi_7$ follows from the previous Lemmas and from the assumed rate conditions. □

Lemma 5. $\|\widehat{\Omega}^{-1}P'\mathcal{M}(g_0(X) - P\beta_{g_0,K})/n\|_2 = O_p(K^{-\alpha_{g_0}})$.

Proof.

$$\begin{aligned}\|\widehat{\Omega}^{-1}P'\mathcal{M}(g_0(X) - P\beta_{g_0,K})/n\|_2 &= [(g_0(X) - P\beta)'\mathcal{M}P\widehat{\Omega}^{-1}P'\mathcal{M}(g_0(X) - P\beta)/n]^{1/2} \\ &= O_p(1)[(g_0(X) - P\beta_{g_0,K})'(g_0(Z) - P\beta_{g_0,K})/n]^{1/2} \\ &= O_p(K^{-\alpha_{g_0}})\end{aligned}$$

by assumption on $(g_0(X) - P\beta_{g_0,K})$ and idempotency of $\mathcal{M}P\widehat{\Omega}^{-1}P'\mathcal{M} = \mathcal{M}P(P'\mathcal{M}\mathcal{M}P)^{-1}\mathcal{M}$. □

Lemma 6. $\|\widehat{\Omega}^{-1}P'\mathcal{M}h_0(Z)/n\|_2 = o_p(n^{-1/2})$.

Proof. $\widehat{\Omega}$ has eigenvalues bounded below and above with probability approaching 1. Then,

$$\begin{aligned}
\|\widehat{\Omega}^{-1}P'\mathcal{M}h_0(Z)/n\|_2 &\leqslant O_p(1)\|P'\mathcal{M}h_0(Z)/n\|_2\\
&= O_p(1)\|(m+W)'\mathcal{M}h_0(Z)/n\|_2\\
&= n^{-1/2}O_p(1)n^{-1/2}\|(m+W)'\mathcal{M}h_0(Z)/n\|_2\\
&\leqslant n^{-1/2}O_p(1)(n^{-1/2}\|m'\mathcal{M}h_0(Z)\|_2 + n^{-1/2}\|W'\mathcal{M}h_0(Z)\|_2)\\
&= n^{-1/2}O_p(1)(\Xi_4+\Xi_5)\\
&= n^{-1/2}O_p(1)o_p(1).
\end{aligned}$$

□

Lemma 7. $\|\widehat{\beta}_g - \beta_{g_0,K}\|_2 = O_p(n^{-1/2}K^{1/2} + K^{-\alpha_{g_0}})$.

Proof. Note that $([\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_g - \beta_{g_0,K}) = n^{-1}\widehat{\Omega}^{-1}P'\mathcal{M}\mathcal{E} + n^{-1}\widehat{\Omega}^{-1}P'M_{\widehat{I}}(g_0(X) - P\beta_{g_0,K}) + n^{-1}\widehat{\Omega}^{-1}P'\mathcal{M}h_0(Z)$. Triangle inequality in conjuction with the bounds described in the previous three lemmas give the result. □

The final statement of Theorem 1 follows from the bound on $\|\widehat{\beta}_g - \beta_{g_0,K}\|_2$ using the arguments in [4]. ■

B.4. **Proof of Theorem 2.** Recall that $F = V^{-1/2}$. Let $\bar{g} = p^K(x)'\beta_{g_0,K}$ and decompose the quantity $n^{1/2}F[a(\widehat{g}) - a(g_0)]$ by

$$\begin{aligned}
n^{1/2}F[a(\widehat{g}) - a(g_0)] = n^{1/2}F[&a(\widehat{g}) - a(g_0) + D(\widehat{g}) - D(g_0)\\
&+ D(\bar{g}) - D(\widehat{g})\\
&+ D(g_0) - D(\bar{g})].
\end{aligned}$$

Lemma 8. $n^{1/2}F[D(\bar{g}) - D(g_0)] = O(n^{1/2}K^{-\alpha_{g_0}})$.

Proof. This follows from arguments given in the proof of Theorem 2 in [4]. Note that the statement does not contain any reference to random quantities. □

Lemma 9. $|n^{1/2}F[a(\widehat{g}) - a(g) - D(\widehat{g}) + D(g)| = o_p(1)$.

Proof. Bounds on $|\widehat{g} - g|_d$ given by Theorem 1 imply that $|n^{1/2}F[a(\widehat{g}) - a(g_0) - D(\widehat{g}) + D(g_0)| \leqslant Cn^{1/2}|\widehat{g} - g_0|_d^2 = O_p(n^{1/2}(n^{-1/2}\zeta_d(K)K^{1/2} + K^{-\alpha_{g_0}})^2) = o_p(1)$. This is again identical to the reasoning given in Theorem 2 in [4], since that references uses only a bound on $|\widehat{g} - g|_d$ to prove the analogous result. □

The last step is to show that $n^{1/2}F[D(\widehat{g}) - D(\bar{g})] \to_d \mathrm{N}(0,1)$.

Lemma 10. $n^{1/2}F[D(\widehat{g}) - D(\bar{g})] \to_d \mathrm{N}(0,1)$.

Proof. Note that $D(\widehat{g})$ can be expanded

$$
\begin{aligned}
D(\widehat{g}) &= D(p^K(x)'[\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_g) = D(p^K(x)'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}Y) \\
&= D(p^K(x)'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}(g_0(X)+h_0(Z)+\mathcal{E})) \\
&= D(p^K(x))'\widehat{\Omega}^{-1}n^{-1}(g_0(X)+h_0(Z)+\mathcal{E}) \\
&= A'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}(g_0(X)+h_0(Z)+\mathcal{E}) \\
&= A'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}g_0(X) + A'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}h_0(Z) + A'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}\mathcal{E}.
\end{aligned}
$$

In addition, $D(\bar{g}) = D(p^K(x)'\beta_{g_0,K}) = A'\beta_{g_0,K}$ gives

$$
\begin{aligned}
n^{1/2}F[D(\widehat{g}) - D(\bar{g})] &= n^{1/2}FA'[\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}g_0(X) - \beta_{g_0,K}] \\
&\quad + n^{1/2}FA'[\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}(h_0(Z)+\mathcal{E})].
\end{aligned}
$$

The above equation gives a decomposition of the right hand side into two terms, which are next bounded separately. Before proceeding, note that the following bounds $\|FA\|_2 = O(1)$, $\|FA'\widehat{\Omega}^{-1}\|_2 = O_p(1)$, $\|FA'\widehat{\Omega}^{-1/2}\|_2 = O_p(1)$, $\|FA'\Omega^{-1}\|_2 = O_p(1)$, $\|FA'\Omega^{-1/2}\|_2 = O(1)$ all hold by arguments in [4]. Consider the first term.

$$
\begin{aligned}
&|n^{1/2}FA'[n^{-1}\widehat{\Omega}^{-1}P'\mathcal{M}g_0(X) - \beta_{g_0,K}]| \\
&= |\sqrt{n}FA'[(P'\mathcal{M}P/n)^{-1}P'\mathcal{M}(G-P\beta)/n]| \\
&\leqslant \|FA'\widehat{\Omega}^{-1}P'\mathcal{M}/\sqrt{n}\|_2\|g_0(X) - P\beta_{g_0,K}\|_2 \\
&\leqslant \|FA'\widehat{\Omega}^{-1}P'\mathcal{M}/\sqrt{n}\|_2\sqrt{n}\max_{i\leqslant n}|g(x_i) - \bar{g}(x_i)| \\
&= \|FA'\widehat{\Omega}^{-1/2}\|_2\sqrt{n}\max_{i\leqslant n}|g(x_i) - \bar{g}(x_i)| \\
&\leqslant \|FA'\widehat{\Omega}^{-1/2}\|_2\sqrt{n}|g-\bar{g}|_0 \\
&= O_p(1)O_p(\sqrt{n}K^{-\alpha}) \\
&= o_p(1).
\end{aligned}
$$

Next, consider $n^{1/2}FA'\widehat{\Omega}^{-1}n^{-1}P'\mathcal{M}(h_0(Z)+\mathcal{E})$. To handle this term, first bound

$$
\begin{aligned}
&|n^{-1/2}FA'(\widehat{\Omega}^{-1} - \Omega^{-1})P'\mathcal{M}(h_0(Z)+\mathcal{E})| \\
&\leqslant n^{-1/2}\|FA'(\widehat{\Omega}^{-1} - \Omega^{-1})\|_2\|P'\mathcal{M}(h_0(Z)+\mathcal{E})\|_2 \\
&= \|FA'(\widehat{\Omega}^{-1} - \Omega^{-1})\|_2(n^{-1/2}\|P'\mathcal{M}(h_0(Z)+\mathcal{E})\|_2) \\
&= \|FA'(\widehat{\Omega}^{-1} - \Omega^{-1})\|_2(n^{-1/2}\|(m+W)'\mathcal{M}(h_0(Z)+\mathcal{E})\|_2) \\
&\leqslant \|FA'(\widehat{\Omega}^{-1} - \Omega^{-1})\|_2(\Xi_4+\Xi_5+\Xi_6+\Xi_7) \\
&\leqslant \|\widehat{\Omega}^{-1} - \Omega^{-1}\|_{2\to2}\|FA'\|_2(\Xi_4+\Xi_5+\Xi_6+\Xi_7) \\
&= o_p(1).
\end{aligned}
$$

Next consider the last remaining term for which a central limit result will be shown.

$$
\begin{aligned}
&\sqrt{n}FA'\Omega^{-1}P'\mathcal{M}(h_0(Z)+\mathcal{E})/n \\
&= \sqrt{n}FA'\Omega^{-1}(W+m)'\mathcal{M}(h_0(Z)+\mathcal{E})/n \\
&= \sqrt{n}(FA'\Omega^{-1}W+m_a)'\mathcal{M}(h_0(Z)+\mathcal{E})/n \\
&= \sqrt{n}FA'\Omega^{-1}W\mathcal{M}\mathcal{E}+\sqrt{n}m_a'\mathcal{M}(h_0(Z)+\mathcal{E})/n+\sqrt{n}W_a'\mathcal{M}h_0(Z)/n \\
&= \sqrt{n}FA'\Omega^{-1}W\mathcal{E}-\sqrt{n}FA'\Omega^{-1}W\mathcal{P}\mathcal{E}+\sqrt{n}m_a'\mathcal{M}(h_0(Z)+\mathcal{E})/n+\sqrt{n}W_a'\mathcal{M}h_0(Z)/n \\
&= \sqrt{n}FA'\Omega^{-1}W'\mathcal{E}/n+o_p(1).
\end{aligned}
$$

Note that the last $o_p(1)$ bound in the equation array above holds by the fact that $|\sqrt{n}FA'\Omega^{-1}W\mathcal{P}\mathcal{E}+\sqrt{n}m_a'\mathcal{M}(h_0(Z)+\mathcal{E})/n+\sqrt{n}W_a'\mathcal{M}h_0(Z)/n| \leqslant \Xi_8+\Xi_9+\Xi_{10}+\Xi_{11}$. The term $\sqrt{n}FA'\Omega^{-1}W'\mathcal{E}/n$ satisfies the conditions Lindbergh-Feller Central Limit Theorem, by arguments given in [4]. □

The previous three lemmas prove that $n^{1/2}F[a(\widehat{g})-a(g_0)] \to \mathrm{N}(0,1)$.

The next set of arguments bound $\widehat{V}-V$. For ν as in the statement of Assumption 14, Define the event $\mathcal{A}_g=\{|\widehat{g}-g_0|_d<\nu/2\}$. Define $\widehat{u}=1_{\mathcal{A}_g}\widehat{\Omega}^{-1}\widehat{A}F$ and $u=1_{\mathcal{A}_g}\Omega^{-1}AF$. In addition, define $\bar{\Sigma}=\sum_i W_iW_i'\varepsilon_i^2/n$, an infeasible sample analogue of Σ.

Lemma 11.

1. $\|\widehat{A}-A\|_2=o_p(1)$
2. $\|\widehat{u}-u\|_2=o_p(1)$
3. $\|\bar{\Sigma}-\Sigma\|_{\mathcal{F}}=o_p(1)$
4. $|\widehat{u}\bar{\Sigma}\widehat{u}-\widehat{u}'\Sigma\widehat{u}|=o_p(1)$.

Proof. **Statement 1.** In the case that $a(g)$ is linear in g, then $a(p'\beta)=A'\beta \implies \widehat{A}=A$. Therefore, consider the case that $a(g)$ is not linear in g. Using arguments identical to those in [4], $1_{\mathcal{A}_g}=1$ with probability $\to 1$, and

$$1_{\mathcal{A}_g}\|\widehat{A}-A\|_2 \leqslant C\cdot\zeta_d(K)|\widehat{g}-g|_d.$$

Statement 2. This follows from arguments in [4].

Statement 3. This follows from arguments in [4].

Statement 4. An immediate implication of Statement 3 is that $1_{\mathcal{A}_g}|\widehat{u}'\bar{\Sigma}\widehat{u}-\widehat{u}'\Sigma\widehat{u}|=|\widehat{u}'(\bar{\Sigma}-\Sigma)\widehat{u}| \leqslant \|\widehat{u}\|_2^2\|\bar{\Sigma}-\Sigma\|_{2\to 2}^2=O_p(1)o_p(1)$.

□

Lemma 12. $\max_{i\leqslant n}|h_0(z_i)-\widehat{h}(z_i)|=o_p(1)$.

Proof. First note that

$$\max_i|h_0(z_i)-\widehat{h}(z_i)|\leqslant\max_i|h_0(z_i)-q^L(z_i)'\beta_{h_0,L,s_0}|$$
$$+\max_i|q^L(z_i)'[\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_h-q^L(z_i)'\beta_{h_0,L,s_0}|.$$

The first term has the bound $\max_i|h_0(z_i)-q(x_i)'\eta|=O_p(L^{-\alpha_z})$ by assumption. Next,

$$\begin{aligned}\max_i|q^L(z_i)'[\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_h-q^L(z_i)'\beta_{h_0,L,s_0}|&=\max_i|q^L(z_i)'([\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_h-\beta_{h_0,L,s_0})|\\&\leqslant\max_i\|q^L(z_i)\|_\infty\|[\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_h-\beta_{h_0,L,s_0}\|_1\end{aligned}$$

Then,

$$\begin{aligned}\|[\widehat{\beta}_{y,(\tilde{p},\tilde{q})}]_h-\beta_{h_0,L,s_0}\|_1&=\|\widehat{\beta}_{y-\widehat{g},I_{\Phi_K+\mathrm{RF}}}-\beta_{h_0,L,s_0}\|_1\\&=\|\widehat{\beta}_{g_0,I_{\Phi_K+\mathrm{RF}}}+\widehat{\beta}_{h_0,I_{\Phi_K+\mathrm{RF}}}+\widehat{\beta}_{\varepsilon,I_{\Phi_K+\mathrm{RF}}}-\widehat{\beta}_{\widehat{g},I_{\Phi_K+\mathrm{RF}}}-\beta_{h_0,L,s_0}\|_1.\\&\leqslant\|\widehat{\beta}_{h_0,I_{\Phi_K+\mathrm{RF}}}-\beta_{h_0,L,s_0}\|_1+\|\widehat{\beta}_{\varepsilon,I_{\Phi_K+\mathrm{RF}}}\|_1+\|\widehat{\beta}_{g_0-\widehat{g},I_{\Phi_K+\mathrm{RF}}}\|_1\\&=J_8+J_{10}+\|\widehat{\beta}_{g_0-\widehat{g},I_{\Phi_K+\mathrm{RF}}}\|_1.\end{aligned}$$

Next,

$$\begin{aligned}&\|\widehat{\beta}_{g_0-\widehat{g},I_{\Phi_K+\mathrm{RF}}}\|_1\\&\leqslant|I_{\Phi_K+\mathrm{RF}}|^{1/2}\|\widehat{\beta}_{g_0-\widehat{g},I_{\Phi_K+\mathrm{RF}}}\|_2\\&=|I_{\Phi_K+\mathrm{RF}}|^{1/2}\|(Q'_{I_{\Phi_K+\mathrm{RF}}}Q_{I_{\Phi_K+\mathrm{RF}}}/n)^{-1}Q'_{I_{\Phi_K+\mathrm{RF}}}(g_0(X)-\widehat{g}(X))/n\|_2\\&\leqslant|I_{\Phi_K+\mathrm{RF}}|^{1/2}\kappa_{\min}(|I_{\Phi_K+\mathrm{RF}}|)^{-1}\|Q'_{I_{\Phi_K+\mathrm{RF}}}(g_0(X)-\widehat{g}(X))/n\|_2\\&\leqslant|I_{\Phi_K+\mathrm{RF}}|^{1/2}\kappa_{\min}(|I_{\Phi_K+\mathrm{RF}}|)^{-1}|I_{\Phi_K+\mathrm{RF}}|^{1/2}\|Q'_{I_{\Phi_K+\mathrm{RF}}}(g_0(X)-\widehat{g}(X))/n\|_\infty\\&\leqslant|I_{\Phi_K+\mathrm{RF}}|\kappa_{\min}(|I_{\Phi_K+\mathrm{RF}}|)^{-1}\|Q'_{I_{\Phi_K+\mathrm{RF}}}(g_0(X)-\widehat{g}(X))/n\|_\infty\\&=|I_{\Phi_K+\mathrm{RF}}|\kappa_{\min}(|I_{\Phi_K+\mathrm{RF}}|)^{-1}\left(\max_j n^{-1}\sum_{i=1}^n|q_{jL}(z_i)|\right)\|g_0(X)-\widehat{g}(X)\|_\infty\\&=O_p(s_0^1K^{\alpha_{I_\Phi}})O_p(1)o_p(n^{-1/2}\zeta_0(K)K^{1/2}+K^{-\alpha_{g_0}}).\end{aligned}$$

Putting these together, it follows from the assumed rate conditions that

$$\max_i|h_0(z_i)-\widehat{h}(z_i)|=o_p(1).$$

□

Next, let $\Delta_{g_0i}=g_0(x_i)-\widehat{g}(x_i)$ and $\Delta_{h_0i}=h_0(z_i)-\widehat{h}(z_i)$. Then above lemma states $\max_{i\leqslant n}\Delta_{h_0i}=o_p(1)$. In addition $\max_{i\leqslant n}|\Delta_{g_0i}|\leqslant|\widehat{g}-g|_0=o_p(1)$. Let $\omega_i^2=u'W_iW_i'u$ and $\widehat{\omega}_i^2=\widehat{u}'W_iW_i'\widehat{u}$.

Lemma 13. $|F\widehat{V}F-\widehat{u}'\bar{\Sigma}\widehat{u}|=o_p(1)$.

Proof.

$$1_{\mathcal{A}_g}|F\widehat{V}F - \widehat{u}'\bar{\Sigma}\widehat{u}| = |\widehat{u}'(\widehat{\Sigma} - \bar{\Sigma})\widehat{u}| = \left|\sum_{i=1}^n \widehat{u}'\widehat{W}_i\widehat{W}_i'\widehat{\varepsilon}_i^2\widehat{u}/n - \sum_{i=1}^n \widehat{u}'W_iW_i'\varepsilon_i^2\widehat{u}/n\right|$$
$$\leqslant \left|\sum_{i=1}^n \omega_i^2(\widehat{\varepsilon}_i^2 - \varepsilon_i^2)/n\right| + \left|\sum_{i=1}^n (\widehat{\omega}_i^2 - \omega_i^2)\widehat{\varepsilon}_i^2/n\right|.$$

Both terms on the right hand side will be bounded. Consider the first term. Expanding $(\widehat{\varepsilon}_i^2 - \varepsilon_i^2)$ gives

$$\left|\sum_{i=1}^n \omega_i^2(\widehat{\varepsilon}_i^2 - \varepsilon_i^2)/n\right| \leqslant \left|\sum_{i=1}^n \omega_i^2\Delta_{1i}^2/n\right| + \left|\sum_{i=1}^n \omega_i^2\Delta_{2i}^2/n\right| + \left|\sum_{i=1}^n \omega_i^2\Delta_{1i}\Delta 2i/n\right|$$
$$+2\left|\sum_{i=1}^n \omega_i^2\Delta_{1i}\varepsilon_i/n\right| + 2\left|\sum_{i=1}^n \omega_i^2\Delta_{2i}\varepsilon_i/n\right|.$$

Note that $\sum_{i=1}^n \omega_i^2/n, \sum_{i=1}^n \omega_i^2|\varepsilon_i| = O_p(1)$ by arguments is in [4]. The five terms above are then bounded in order of their appearence by

$$\sum_{i=1}^n \omega_i^2\Delta_{1i}^2/n \leqslant \max_{i\leqslant n}|\Delta_{1i}|\sum_{i=1}^n \omega_i^2/n = o_p(1)O_p(1)$$
$$\sum_{i=1}^n \omega_i^2\Delta_{2i}^2/n \leqslant \max_{i\leqslant n}|\Delta_{2i}|\sum_{i=1}^n \omega_i^2|\varepsilon_i|/n = o_p(1)O_p(1)$$
$$\sum_{i=1}^n \omega_i^2\Delta_{1i}\Delta_{2i}/n \leqslant \max_{i\leqslant n}|\Delta_{1i}|\max_{i\leqslant n}|\Delta_{2i}|\sum_{i=1}^n \omega_i^2/n = o_p(1)O_p(1)$$
$$\sum_{i=1}^n \omega_i^2\Delta_{1i}\varepsilon_i/n \leqslant \max_{i\leqslant n}|\Delta_{1i}|\sum_{i=1}^n \omega_i^2|\varepsilon_i|/n = o_p(1)O_p(1)$$
$$\sum_{i=1}^n \omega_i^2\Delta_{2i}\varepsilon_i/n \leqslant \max_{i\leqslant n}|\Delta_{2i}|\sum_{i=1}^n \omega_i^2|\varepsilon_i|/n = o_p(1)O_p(1).$$

The second term is bounded by

$$\left|\sum_{i=1}^n \widehat{u}'(\widehat{W}_i\widehat{W}_i' - W_iW_i')\widehat{\varepsilon}_i^2\widehat{u}/n\right| \leqslant \max_{i\leqslant n}|\varepsilon_i^2|\left|\sum_{i=1}^n \widehat{u}'(\widehat{W}_i\widehat{W}_i' - W_iW_i')\widehat{u}/n\right|$$
$$\leqslant \max_{i\leqslant n}|\widehat{\varepsilon}_i^2|\|\widehat{u}\|_2^2\|\sum_{i=1}^n(\widehat{W}_i\widehat{W}_i' - W_iW_i')/n\|_{2\to 2} = \max_{i\leqslant n}|\widehat{\varepsilon}_i^2|\|\widehat{u}\|_2^2\|\|\widehat{\Omega} - \bar{\Omega}\|_{2\to 2}$$
$$\leqslant \left(\max_{i\leqslant n}|\varepsilon_i^2| + \max_{i\leqslant n}|\widehat{\varepsilon}_i^2 - \varepsilon_i^2|\right)\|\widehat{u}\|_2^2\|\|\widehat{\Omega} - \bar{\Omega}\|_{2\to 2}$$
$$= \left(O_p(n^{2/\delta}) + o_p(1)\right)O_p(1)(O_p(\zeta_0(K)n^{-1/2}K^{1/2}) + \Xi_1 + \Xi_2 + \Xi_3)$$
$$= o_p(1).$$

where the last bounds come from the rate condition in Assumption 9 and $\max_{i\leqslant n}|\widehat{\varepsilon}_i^2 - \varepsilon_i^2| = o_p(1)$ by $\max_{i\leqslant n}|\Delta_{1i}| + |\Delta_{2i}| = o_p(1)$. $\square$

These results give the conclusion that

$$n^{1/2}\widehat{V}^{-1/2}(\widehat{\theta} - \theta) = n^{1/2}(F\widehat{V}F)^{-1/2}(\widehat{\theta} - \theta) \xrightarrow{d} N(0,1).$$

Calculations which give the rates of convergence in each of the cases of Assumption 17 or of Assumption 18, as well as the proof of the second statement of Theorem 2, use the same arguments as in [4]. This concludes the proof. ■

References

[1] A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80:2369–2429, 2012. Arxiv, 2010.
[2] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection amongst high-dimensional controls with an application to abortion on crime. *Review of Economic Studies*, 81(2):608–650, 2014.
[3] C. Hansen, D. Kozbur, and S. Misra. Targeted Undersmoothing. *ArXiv e-prints*, June 2017.
[4] Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.