SUPPLEMENT TO "CONSTRUCTING OPTIMAL INSTRUMENTS BY FIRST-STAGE PREDICTION AVERAGING": AUXILIARY APPENDIX

(*Econometrica*, Vol. 78, No. 2, March 2010, 697–718)

BY GUIDO KUERSTEINER AND RYO OKUI

This appendix contains detailed proofs for the results given in the main paper. For ease of reference, we repeat formal assumptions, some key definitions, and statements of propositions and theorems.

1. DEFINITIONS

FOR EASE OF REFERENCE, we repeat the definitions and formulas that appear in the main text. Enumerations that appear only in this document are referenced by numbers preceded by "A." The definitions of the estimators are repeated first. The model averaging two-stage least squares (MA2SLS) estimator is

$$\hat{\beta} = (X'P(W)X)^{-1}X'P(W)y. \tag{2.2}$$

The definition of (2.2) can be extended to the LIML estimator. Let

$$\hat{\Lambda}_m = \min_{\beta} \frac{(y - X\beta)'P_m(y - X\beta)}{(y - X\beta)'(y - X\beta)}$$

and define $\hat{\Lambda}(W) = \sum_{m=1}^{M} w_m \hat{\Lambda}_m$. The MALIML estimator, $\hat{\beta}$, of β then is defined as

$$\hat{\beta} = (X'P(W)X - \hat{\Lambda}(W)X'X)^{-1}(X'P(W)y - \hat{\Lambda}(W)X'y). \tag{2.3}$$

Similarly we consider a modification to Fuller's (1977) estimator. Let

$$\check{\Lambda}_m = \left(\frac{\hat{\Lambda}_m - \dfrac{\alpha}{N-m}(1 - \hat{\Lambda}_m)}{1 - \dfrac{\alpha}{N-m}(1 - \hat{\Lambda}_m)} \right),$$

where α is a constant. The model averaging Fuller estimator (MAFuller) then is defined as

$$\hat{\beta} = (X'P(W)X - \check{\Lambda}(W)X'X)^{-1}(X'P(W)y - \check{\Lambda}(W)X'y). \tag{2.4}$$

The choice of model weights W is based on an approximation to the higher order MSE of $\hat{\beta}$. Following Donald and Newey (2001) (see also Nagar (1959)), we approximate the MSE conditional on the exogenous variable z, $E[(\hat{\beta} -$

 DOI: 10.3982/ECTA7444

$\beta_0)(\hat\beta-\beta_0)'|z]$, by $\sigma_\epsilon^2 H^{-1}+S(W)$, where

$$N(\hat\beta-\beta_0)(\hat\beta-\beta_0)'=\hat Q(W)+\hat r(W),\qquad E[\hat Q(W)|z]=\sigma_\epsilon^2 H^{-1}+S(W)+T(W), \tag{4.1}$$

$H=f'f/N$, and $(\hat r(W)+T(W))/\operatorname{tr}(S(W))=o_p(1)$ as $N\to\infty$. However, because of the possibility of bias elimination by setting $K'W=0$, we need to consider an expansion that contains additional higher order terms for the MA2SLS case. We show the asymptotic properties of the MA2SLS, MALIML, and MAFuller estimators under the assumptions in Section A.

Next we discuss the estimation of $S(W)$. Let $\tilde\beta$ denote some preliminary estimator of β and define the residuals $\tilde\epsilon=y-X\tilde\beta$. As pointed out in Donald and Newey (2001), it is important that $\tilde\beta$ does not depend on the weighting vector, W. We use the 2SLS estimator with the number of instruments selected by the first-stage Mallows criterion in simulations for MA2SLS, and use the corresponding LIML and Fuller estimator for MALIML and MAFuller. Let $\hat H$ be some estimator of $H=f'f/n$. Let $\tilde u$ be some preliminary residual vector of the first-stage regression. Let $\tilde u_\lambda=\tilde u\hat H^{-1}\lambda$. Define

$$\hat\sigma_\epsilon^2=\tilde\epsilon'\tilde\epsilon/N,\quad \hat\sigma_\lambda^2=\tilde u_\lambda'\tilde u_\lambda/N,\quad \hat\sigma_{\lambda\epsilon}=\tilde u_\lambda'\tilde\epsilon/N.$$

Let $\hat u_\lambda^m=(P_M-P_m)X\hat H^{-1}\lambda$ and $\hat U=(\hat u_\lambda^1,\dots,\hat u_\lambda^M)'(\hat u_\lambda^1,\dots,\hat u_\lambda^M)$.[1] Let Γ be the $M\times M$ matrix whose (i,j) element is $\min(i,j)$ and let $K=(1,2,\dots,M)'$. The criterion $\hat S_\lambda(W)$ for choosing the weights for MA2SLS is

$$\hat S_\lambda(W)=\hat a_\lambda\frac{(K'W)^2}{N}+\hat b_\lambda\frac{(W'\Gamma W)}{N}-\frac{K'W}{N}\hat B_{\lambda,N}+\hat\sigma_\epsilon^2\left(\frac{W'\hat UW-\hat\sigma_\lambda^2(M-2K'W+W'\Gamma W)}{N}\right) \tag{2.5}$$

with $\hat a_\lambda=\hat\sigma_{\lambda\epsilon}^2$, $\hat b_\lambda=\hat\sigma_\epsilon^2\hat\sigma_\lambda^2+\hat\sigma_{\lambda\epsilon}^2$, and $\hat B_{\lambda,N}=\lambda'\hat H^{-1}\hat B_N\hat H^{-1}\lambda$, where $\hat B_N$ can be estimated by

$$\hat B_N=2\left(\hat\sigma_\epsilon^2\hat\Sigma_u+d\hat\sigma_{u\epsilon}\hat\sigma_{u\epsilon}'+\frac{1}{N}\sum_{i=1}^N\hat f_i\hat\sigma_{u\epsilon}'\hat H^{-1}\hat\sigma_{u\epsilon}\hat f_i'+\frac{1}{N}\sum_{i=1}^N(\hat f_i\hat\sigma_{u\epsilon}'\hat H^{-1}\hat f_i\hat\sigma_{u\epsilon}'+\hat\sigma_{u\epsilon}\hat f_i'\hat H^{-1}\hat\sigma_{u\epsilon}\hat f_i')\right)$$

[1] Note that $\tilde u$ is the preliminary residual vector, but $\hat u_\lambda^m$s are the vectors of the differences of the residuals.

in which $\hat{f} = P_m X$ with $m \to \infty$ (m may be chosen by the first-stage Mallows criterion) and $\hat{\sigma}_{u\epsilon} = \tilde{u}'\tilde{\epsilon}/N$. When the weights are only allowed to be positive, we can use the simpler criterion

$$\hat{S}_\lambda(W) = \hat{a}_\lambda \frac{(K'W)^2}{N} + \hat{\sigma}_\epsilon^2 \left(\frac{W'\hat{U}W - \hat{\sigma}_\lambda^2(M - 2K'W + W'\Gamma W)}{N} \right) \tag{2.6}$$

that does not account for the smaller orders terms involving $\hat{b}_\lambda$ and $\hat{B}_{\lambda,N}$. For MALIML and MAFuller, we choose W based on the criterion

$$\begin{aligned} \hat{S}_\lambda(W) = {} & (\hat{\sigma}_\epsilon^2 \hat{\sigma}_\lambda^2 - \hat{\sigma}_{\lambda\epsilon}^2) \frac{W'\Gamma W}{N} \\ & + \hat{\sigma}_\epsilon^2 \left(\frac{W'\hat{U}W - \hat{\sigma}_\lambda^2(M - 2K'W + W'\Gamma W)}{N} \right). \end{aligned} \tag{2.7}$$

A. REGULARITY CONDITIONS AND FORMAL RESULTS

ASSUMPTION 1: *$\{y_i, X_i, z_i\}$ are i.i.d., $E[\epsilon_i^2|z_i] = \sigma_\epsilon^2 > 0$, and $E[\|\eta_i\|^4|z_i]$ and $E[|\epsilon_i|^4|z_i]$ are bounded.*

ASSUMPTION 2: (i) *$\bar{H} \equiv E[f_i f_i']$ exists and is nonsingular.* (ii) *for some $\alpha > 1/2$,*

$$\sup_{m \le M} m^{2\alpha} \Big(\sup_{\lambda'\lambda = 1} \lambda' f'(I - P_m) f \lambda / N \Big) = O_p(1).$$

(iii) *Let N_+ be the set of positive integers. There exists a subset $\bar{J} \subset N_+$ with a finite number of elements such that $\sup_{m \in \bar{J}} \sup_{\lambda'\lambda} \lambda' f'(P_m - P_{m+1}) f \lambda / N = 0$ wpa1 and for all $m \notin \bar{J}$, it follows that*

$$\inf_{m \notin \bar{J}, m \le M} m^{2\alpha+1} \Big(\sup_{\lambda'\lambda = 1} \lambda' f'(P_m - P_{m+1}) f \lambda / N \Big) > 0 \quad wpa1.$$

ASSUMPTION 3: (i) *Let u_{ia} be the ath element of u_i. Then $E[\epsilon_i^r u_{ia}^s|z_i]$ are constant and bounded for all a and $r, s \ge 0$ and $r + s \le 5$. Let $\sigma_{u\epsilon} = E[u_i \epsilon_i|z_i]$ and $\Sigma_u = E[u_i u_i'|z_i]$.* (ii) *$Z_M' Z_M$ are nonsingular wpa1.* (iii) *$\max_{i \le N} P_{M,ii} \to_p 0$, where $P_{M,ii}$ signifies the (i,i)th element of P_M.* (iv) *f_i is bounded.*

ASSUMPTION 4: *Let $W^+ = (|w_{1,N}|, \ldots, |w_{M,N}|)'$. The following conditions hold: $1_M' W = 1$; $W \in l_1$ for all N where $l_1 = \{x = (x_1, \ldots) | \sum_{i=1}^\infty |x_i| \le C_{l1} < \infty\}$ for some constant C_{l1}, $M \le N$; and, as $N \to \infty$ and $M \to \infty$, $K'W^+ = \sum_{m=1}^M |w_m| m \to \infty$. For some sequence $L \le M$ such that $L \to \infty$*

as $N \to \infty$ *and* $L \notin \bar{J}$, *where* $\bar{J}$ *is defined in Assumption* 2(iii) *it follows that* $\sup_{j \notin \bar{J}, j \le L} |\sum_{m=1}^{j} w_m| = O(1/\sqrt{N})$ *as* $N \to \infty$.

ASSUMPTION 5: *It holds either that* (i) $K'W^+/\sqrt{N} = \sum_{m=1}^{M} |w_m| m/\sqrt{N} \to 0$ *or* (ii) $K'W^+/N = \sum_{m=1}^{M} |w_m| m/N \to 0$ *and* $M/N \to 0$.

ASSUMPTION 6: *The eigenvalues of* $E[Z_{k,i}Z'_{k,i}]$ *are bounded away from zero uniformly in* k. *Let* $\bar{H}_k = E[f_i Z_{k,i}](E[Z_{k,i}Z'_{k,i}])^{-1}E[f_i Z'_{k,i}]'$ *and* $\bar{H} = E[f_i f'_i]$. *Then* $\|\bar{H}_k - \bar{H}\| = O(k^{-2\alpha})$ *for* $k \to \infty$. $E[|\epsilon_i|^8|z]$ *and* $E[|u_{ia}|^8|z]$ *are uniformly bounded in* z *for all* a.

ASSUMPTION 7: $\beta \in \Theta$, *where* Θ *is a compact subset of* R^d.

ASSUMPTION 8: *For some* α, $\sup_{m \le M} m^{2\alpha+1}(\sup_{\lambda'\lambda=1} \lambda' f'(P_m - P_{m+1}) f \lambda/N) = O_p(1)$.

ASSUMPTION 9: $\hat{H} - H = o_p(1)$, $\hat{\sigma}^2_\epsilon - \sigma^2_\epsilon = o_p(1)$, $\hat{\sigma}^2_\lambda - \sigma^2_\lambda = o_p(1)$, $\hat{\sigma}_{\lambda\epsilon} - \sigma_{\lambda\epsilon} = o_p(1)$, *and* $\hat{B}_N - B_N = o_p(1)$.

ASSUMPTION 10: *Let* α *be as defined in Assumption* 8. *For some* $0 < \varepsilon < \min(1/(2\alpha), 1)$ *and* δ *such that* $2\alpha\varepsilon > \delta > 0$, *it holds that* $M = O(N^{(1+\delta)/(2\alpha+1)})$. *For some* $\vartheta > (1+\delta)/(1-2\alpha\varepsilon)$, *it holds that* $E(|u_i|^{2\vartheta}) < \infty$. *Further assume that* $\hat{\sigma}^2_\lambda - \sigma^2_\lambda = o_p(N^{-\delta/(2\alpha+1)})$.

THEOREM A.1: *Suppose that Assumptions* 1–3 *are satisfied. Define* $\mu_i(W) = E[\epsilon_i^2 u_i] P_{ii}(W)$ *and* $\mu(W) = (\mu_1(W), \ldots, \mu_N(W))'$. *If* W *satisfies Assumptions* 4 *and* 5(i), *then, for* $\hat{\beta}$ *defined in* (2.2) (*MA2SLS*), *the decomposition given by* (A.8) *holds with*

$$
\begin{aligned}
S(W) = H^{-1}\Bigg(&\mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u'_i]\frac{\sum_{i=1}^{N}(P_{ii}(W))^2}{N} + \sigma_{u\epsilon}\sigma'_{u\epsilon}\frac{(K'W)^2}{N} \\
&+ (\sigma^2_\epsilon \Sigma_u + \sigma_{u\epsilon}\sigma'_{u\epsilon})\frac{(W'\Gamma W)}{N} \\
&- \frac{K'W}{N}B_N + E[\epsilon_1^2 u_1]\frac{\sum_{i=1}^{N} f'_i P_{ii}(W)}{N} + \frac{\sum_{i=1}^{N} f_i P_{ii}(W)}{N} E[\epsilon_1^2 u'_1] \\
&+ \frac{f'(I - P(W))\mu(W)}{N} + \frac{\mu(W)'(I - P(W))f}{N} \\
&+ \sigma^2_\epsilon \frac{f'(I - P(W))(I - P(W))f}{N}\Bigg)H^{-1},
\end{aligned}
$$

where $d = \dim(\beta)$ *and*

$$\text{(A.1)} \qquad B_N = 2\Bigg(\sigma_\epsilon^2 \Sigma_u + d\sigma_{u\epsilon}\sigma_{u\epsilon}' + \frac{1}{N}\sum_{i=1}^{N} f_i \sigma_{u\epsilon}' H^{-1}\sigma_{u\epsilon} f_i' + \frac{1}{N}\sum_{i=1}^{N}(f_i\sigma_{u\epsilon}' H^{-1} f_i \sigma_{u\epsilon}' + \sigma_{u\epsilon} f_i' H^{-1}\sigma_{u\epsilon} f_i')\Bigg).$$

A number of special cases lead to simplifications of the above result. If $\mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] = 0$ and $E[\epsilon_i^2 u_i] = 0$, as would be the case if ϵ_i and u_i were jointly Gaussian, the following result is obtained.

COROLLARY A.1: *Suppose that the same conditions as in Theorem* A.1 *hold, and that in addition* $\mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] = 0$ *and* $E[\epsilon_i^2 u_i] = 0$. *Then, for* $\hat{\beta}$ *defined in* (2.2) (*MA2SLS*), *the decomposition given by* (A.8) *holds with*

$$\text{(A.2)} \qquad S(W) = H^{-1}\Bigg(\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(K'W)^2}{N} + (\sigma_\epsilon^2\Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')\frac{(W'\Gamma W)}{N} - \frac{K'W}{N}B_N + \sigma_\epsilon^2\frac{f'(I-P(W))(I-P(W))f}{N}\Bigg)H^{-1},$$

where B_N *is as defined before*.

Another interesting case arises when W is constrained such that $w_m \in [0, 1]$. We have the following result.

COROLLARY A.2: *Suppose that the same conditions as in Theorem* A.1 *hold and that in addition* $w_m \in [0, 1]$ *for all* m. *Then, for* $\hat{\beta}$ *defined in* (2.2) (*MA2SLS*), *the decomposition given by* (A.8) *holds with*

$$\text{(A.3)} \qquad S(W) = H^{-1}\Bigg(\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(K'W)^2}{N} + (\sigma_\epsilon^2\Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')\frac{(W'\Gamma W)}{N} - \frac{K'W}{N}B_N + E[\epsilon_1^2 u_1]\frac{\sum_{i=1}^{N} f_i' P_{ii}(W)}{N} + \frac{\sum_{i=1}^{N} f_i P_{ii}(W)}{N}E[\epsilon_1^2 u_1'] + \sigma_\epsilon^2\frac{f'(I-P(W))(I-P(W))f}{N}\Bigg)H^{-1},$$

where B_N is as defined before. Moreover, ignoring terms of order $O_p(K'W)$ $(= o_p((K'W)^2))$, to first order

$$S(W) = H^{-1}\left(\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{(K'W)^2}{N} + \sigma_\epsilon^2\frac{f'(I-P(W))(I-P(W))f}{N}\right)H^{-1}. \tag{A.4}$$

A last special case arises when the constraint $K'W = 0$ is imposed on the weights.

COROLLARY A.3: *Suppose that the same conditions as in Theorem* A.1 *hold and that in addition* $\mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] = 0$ *and* $E[\epsilon_i^2 u_i] = 0$. *Furthermore, impose* $K'W = 0$. *Then, for* $\hat{\beta}$ *defined in* (2.2) (*MA2SLS*), *the decomposition given by* (A.8) *holds with*

$$\begin{aligned} S(W) = H^{-1}\Bigg(&(\sigma_\epsilon^2\Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')\frac{(W'\Gamma W)}{N} \\ &+ \sigma_\epsilon^2\frac{f'(I-P(W))(I-P(W))f}{N}\Bigg)H^{-1}. \end{aligned} \tag{A.5}$$

REMARK A.1: We note that this result covers the Nagar (1959) estimator, where $M = N$, $w_m = N/(N-k)$ for $k = m$, $w_N = -k/(N-k)$, and $w_m = 0$ otherwise for some k such that $k \to \infty$ and $k/\sqrt{N} \to 0$. First, we verify that all the conditions of the corollary are satisfied, where $\sum_{m=1}^M |w_m| = (N + k)/(N-k)$, which is uniformly bounded if $k = o(N)$, $K'W = 1$, $\mathbf{1}_M'W = 1$, $\sum_{m=1}^M |w_m| m = 2Nk/(N-k) \to \infty$, and $\sum_{m=1}^M |w_m| m/\sqrt{N} = 2\sqrt{N}k/(N - k) \to 0$. Further, $\sup_{j \notin \bar{J}, j \le L} |\sum_{m=1}^j w_m| = 0$ by taking $L \le k$. Next, note that $W'\Gamma W = k/(1-k/N)^2 - k^2/N(1-k/N)^2$ and $f'(I-P(W))(I-P(W))f = f'(I-P_k)f/(1-k/N)^2$, noting that $P_N = I$. If we use W_N to denote the Nagar weights, then $S(W_N) = H^{-1}((\sigma_\epsilon^2\Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')k/N + \sigma_\epsilon^2 f'(I-P_k)f/N)H^{-1} + o(S(W_N))$. The lead term is the same as the result in Proposition 3 of Donald and Newey (2001).

The next theorem gives the approximate MSE of the MALIML and MAFuller estimators.

THEOREM A.2: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied. Let* $v_i = u_i - (\sigma_{u\epsilon}/\sigma_\epsilon^2)\epsilon_i$. *Define* $\Sigma_v = \Sigma_u - \sigma_{u\varepsilon}\sigma_{u\varepsilon}'$, $\mu_v(W) = (\mu_{v,1}(W), \ldots, \mu_{v,N}(W))'$, *and* $\mu_{v,i}(W) = E[\epsilon_i^2 v_i]P_{ii}(W)$. *If* W *satisfies Assumption* 4, *then, for* $\hat{\beta}$ *defined in* (2.3) (*MALIML*) *and* $\hat{\beta}$ *defined in* (2.4) (*MAFuller*), *the de-*

composition given by (A.8) *holds with*

$$S(W) = H^{-1}\Bigg(\sigma_\epsilon^2 \Sigma_v \frac{W'\Gamma W}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N} + \mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, v_i'] \frac{\sum_{i=1}^{N}(P_{ii}(W))^2}{N} + \hat{\zeta} + \hat{\zeta}' - \frac{f'(I-P(W))\mu_v(W)}{N} - \frac{\mu_v(W)'(I-P(W))f}{N}\Bigg)H^{-1},$$

where

$$\hat{\zeta} = \sum_{i=1}^{N} f_i P_{ii}(W) E[\epsilon_i^2 v_i']/N - \frac{K'W}{N} \sum_{i=1}^{N} f_i E[\epsilon_i^2 v_i']/N.$$

When $\mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, v_i'] = 0$ *and* $E[\epsilon_i^2 v_i] = 0$, *we have*

$$S(W) = H^{-1}\left(\sigma_\epsilon^2 \Sigma_v \frac{W'\Gamma W}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}\right)H^{-1}. \tag{A.6}$$

The following theorem demonstrates that model averaging can achieve a smaller MSE than that achieved by sequential instrument selection. Define the sets $\Omega_U = \{W \in l_1 | W'\mathbf{1}_M = 1\}$, $\Omega_C = \{W \in l_1 | W'\mathbf{1}_M = 1; w_m \in [-1,1], \forall m \le M\}$, $\Omega_P = \{W \in l_1 | W'\mathbf{1}_M = 1; w_m \in [0,1], \forall m \le M\}$, and $\Omega_B = \{W \in l_1 | W'\mathbf{1}_M = 1, K'W = 0\}$.

THEOREM A.3: *Assume that Assumptions* 1–5 *hold. Let* $\gamma_m = \lambda' H^{-1} \times f'(I - P_m) f H^{-1}\lambda/N$. *Assume that there exists a nonstochastic function* $C(a)$ *such that* $\sup_{a\in[-\varepsilon,\varepsilon]} \gamma_{m(1+a)}/\gamma_m = C(a)$ *wpa*1 *as* $N, m \to \infty$ *for some* $\varepsilon > 0$. *Assume that* $C(a) = (1+a)^{-2\alpha} + o(|a|^{2\alpha})$.

(i) *For* $S_\lambda(W)$ *given by* (A.2), *it follows that*

$$\frac{\min_{W\in\Omega_P} S_\lambda(W)}{\min_{W\in\Omega_{sq}} S_\lambda(W)} < 1 \quad wpa1.$$

Letting W_N be the weights with $w_m = N/(N-m)$, $w_N = -m/(N-m)$, and $w_j = 0$ for $j \neq m$, where m is chosen to minimize $S_\lambda(W)$, it follows that

$$\frac{\min_{W\in\Omega_B} S_\lambda(W)}{S_\lambda(W_N)} < 1 \quad wpa1.$$

(ii) *For $S_\lambda(W)$ given by* (A.6), *it follows that*

$$\frac{\min_{W\in\Omega_P} S_\lambda(W)}{\min_{W\in\Omega_{sq}} S_\lambda(W)} < 1 \quad wpa1.$$

The next theorem is about the optimality of the estimated criterion function. It is a generalization of the result established by Li (1987).

THEOREM A.4: *Let Assumptions* 1–10 *hold. For $\Omega = \Omega_U$, Ω_B, Ω_C, or Ω_P and $\hat{W} = \arg\min_{W\in\Omega} \hat{S}_\lambda(W)$, where $\hat{S}_\lambda(W)$ is defined in either* (2.5) *or* (2.7), *it follows that*

$$\frac{\hat{S}_\lambda(\hat{W})}{\inf_{W\in\Omega} S_\lambda(W)} \to_p 1. \tag{A.7}$$

Last, we present the theorem that shows that the pseudo R^2 converges in probability to the population R^2.

THEOREM A.5: *Assume that Assumptions* 1–5 *hold. Suppose that* $\dim(\beta) = 1$. *Let*

$$\tilde{R}^2 = \frac{(X'P(W)X)^2}{X'P(W)P(W)X \cdot X'X}.$$

If $\sum_{j=1}^{L} |w_j| = o(1)$ and $E(X_i) = 0$, then

$$\tilde{R}^2 \to_p \frac{E(f_i^2)}{E(f_i^2) + \sigma_u^2}.$$

A.1. *Lemmas*

The MA2SLS estimator has the form of $\sqrt{N}(\hat{\beta} - \beta) = \hat{H}^{-1}\hat{h}$. We define $h = f'\epsilon/\sqrt{N}$ and $H = f'f/N$. The following Lemma A.1 is the key device to compute the Nagar-type MSE of MA2SLS. This lemma is similar to Lemma A.1 in Donald and Newey (2001), but with the important difference that the expansion is valid to higher order and covers the case of higher order unbiased estimators.

LEMMA A.1: *If there is a decomposition* $\hat{h} = h + T^h + Z^h$, $\tilde{h} = h + T^h$, $\hat{H} = H + T^H + Z^H$, *and*

$$\tilde{h}\tilde{h}' - \tilde{h}\tilde{h}'H^{-1}T^{H\prime} - T^H H^{-1}\tilde{h}\tilde{h}' = \hat{A}(W) + Z^A(W),$$

such that $T^h = o_p(1)$, $h = O_p(1)$, *and* $H = O_p(1)$, *the determinant of* H *is bounded away from zero with probability* 1, $\rho_{W,N} = \operatorname{tr}(S(W))$, *and* $\rho_{W,N} = o_p(1)$,

$$\|T^H\|^2 = o_p(\rho_{W,N}), \quad \|Z^h\| = o_p(\rho_{W,N}), \quad \|Z^H\| = o_p(\rho_{W,N}),$$
$$Z^A(W) = o_p(\rho_{W,N}), \quad E[\hat{A}(W)|z] = \sigma^2 H + HS(W)H + o_p(\rho_{W,N}),$$

then

$$\begin{aligned} &N(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)' = \hat{Q}(W) + \hat{r}(W), \\ &E[\hat{Q}(W)|z] = \sigma_\epsilon^2 H^{-1} + S(W) + T(W), \\ &(\hat{r}(W) + T(W))/\operatorname{tr}(S(W)) = o_p(1), \quad \text{as} \quad K'W^+ \to \infty, N \to \infty. \end{aligned} \tag{A.8}$$

REMARK A.2: The technical difference between our lemma and that of Donald and Newey is that we consider the interaction between T^h and T^H in the expansion and we do not require that $\|T^h\| \cdot \|T^H\|$ is small.

PROOF OF LEMMA A.1: The proof follows steps taken by Donald and Newey (2001). We observe that

$$\hat{H}^{-1}\hat{h} = H^{-1}\hat{h} - H^{-1}(\hat{H} - H)H^{-1}\hat{h} + H^{-1}(\hat{H} - H)H^{-1}(\hat{H} - H)\hat{H}\hat{h}.$$

Noting that $\hat{H} - H = T^H + Z^H$, $\|T^H\|^2 = o_p(\rho_{W,N})$, $\|Z^H\| = o_p(\rho_{W,N})$, and $\hat{h} = \tilde{h} + Z^h = \tilde{h} + o_p(\rho_{W,N})$, we have

$$\hat{H}^{-1}\hat{h} = H^{-1}\tilde{h} - H^{-1}T^H H^{-1}\tilde{h} + o_p(\rho_{W,N}).$$

Let $\tilde{\tau} = \tilde{h} - T^H H^{-1}\tilde{h}$. Then

$$\begin{aligned} \tilde{\tau}\tilde{\tau}' &= \hat{A}(W) + Z^A(W) + T^H H^{-1}\tilde{h}\tilde{h}'H^{-1}T^H \\ &= \hat{A}(W) + o_p(\rho_{W,N}) \end{aligned}$$

by $Z^A(W) = o_p(\rho_{W,N})$ and $\|T^H\| = o_p(\rho_{W,N})$. It follows that

$$\begin{aligned} N(\hat{\beta} - \beta)(\hat{\beta} - \beta)' &= H^{-1}(\hat{A}(W) + o_p(\rho_{W,N}))H^{-1} + o_p(\rho_{W,N}) \\ &= H^{-1}\hat{A}(W)H^{-1} + o_p(\rho_{W,N}). \end{aligned}$$

Therefore, we get the desired result. *Q.E.D.*

LEMMA A.2: *Let Γ be the $N \times N$ matrix where $\Gamma_{ij} = \min(i, j)$. Then Γ is positive definite.*

PROOF: Define the vectors $b_{j,N} = (\mathbf{0}'_j, \mathbf{1}'_{N-j})'$, where $\mathbf{1}_j$ is the $j \times 1$ vector of 1s and $\mathbf{0}_j$ is defined similarly. Then

$$\Gamma = \sum_{j=0}^{N-1} b_{j,N} b'_{j,N},$$

and for any $y \in \mathbb{R}^N$ it follows that $y'\Gamma y = \sum_{j=0}^{N-1} (y' b_{j,N})^2 \geq 0$ and the equality holds if and only if $y = 0$. This shows that Γ is positive definite. *Q.E.D.*

LEMMA A.3: *Let Γ be defined as in Lemma* A.2. *If, for some sequence $L \leq M$, $L \to \infty$, $L \notin \bar{J}$ for $\bar{J}$ defined in Assumption* 2(iii), $\sup_{j \notin \bar{J}, j \leq L} |\sum_{m=1}^{j} w_m| = O_p(1/\sqrt{N})$ *as $M \to \infty$, and $W'\mathbf{1}_M = 1$ for any M, then it follows that $W'\Gamma W \to \infty$ as $M \to \infty$.*

PROOF: For $L \leq M$ and $L \to \infty$ it follows by the assumption that

$$\begin{aligned} 1 = \left| \sum_{m=1}^{M} w_m \right| &\leq \inf_{j \notin \bar{J}, j \leq L} \left(\left| \sum_{m=j+1}^{M} w_m \right| + \left| \sum_{m=1}^{j} w_m \right| \right) \\ &\leq \inf_{j \notin \bar{J}, j \leq L} \left| \sum_{m=j+1}^{M} w_m \right| + \sup_{j \notin \bar{J}, j \leq L} \left| \sum_{m=1}^{j} w_m \right| \end{aligned}$$

such that $\inf_{j \notin \bar{J}, j \leq L} |\sum_{m=j+1}^{M} w_m| \geq 1 - O_p(1/\sqrt{N})$. Now let $C_{\bar{J}}$ be the number of elements in $\bar{J}$ such that

$$\begin{aligned} W'\Gamma W = \sum_{j=0}^{M-1} \left(\sum_{m=j+1}^{M} w_m \right)^2 &\geq \sum_{j \notin \bar{J}, j \leq L} \left(\sum_{m=j+1}^{M} w_m \right)^2 \\ &\geq (L - C_{\bar{J}})(1 - O_p(1/\sqrt{N}))^2. \end{aligned}$$

Since $L \to \infty$ and $C_{\bar{J}}$ is bounded and does not depend on L or N, the result follows. *Q.E.D.*

LEMMA A.4: *If, for some sequence $L \leq M$, $L \to \infty$, for $\bar{J}$ defined in Assumption* 2(iii), *$L \notin \bar{J}$, and* $\sup_{j \notin \bar{J}, j \leq L} |\sum_{m=1}^{j} w_m| = O(1/\sqrt{N})$ *as $M \to \infty$, then* $\sum_{m=1, m \notin \bar{J}}^{M} (\sum_{s=1}^{m} w_m)^2 m^{-2\alpha} \to 0$.

PROOF: Note that

$$\sum_{m=1,m\notin\bar{J}}^{M}\left(\sum_{s=1}^{m} w_m\right)^2 m^{-2\alpha}$$

$$= \sum_{m=1,m\notin\bar{J}}^{L}\left(\sum_{s=1}^{m} w_m\right)^2 m^{-2\alpha} + \sum_{m=L+1,m\notin\bar{J}}^{M}\left(\sum_{s=1}^{m} w_m\right)^2 m^{-2\alpha}$$

$$\le \left(\sup_{j\notin\bar{J},j\le L}\left|\sum_{s=1}^{j} w_m\right|\right)^2 \sum_{m=1}^{L} m^{-2\alpha} + \sum_{m=L+1,m\notin\bar{J}}^{M}\left(\sum_{s=1}^{m} |w_m|\right)^2 m^{-2\alpha}$$

$$\le O(1/N)\sum_{m=1}^{L} m^{-2\alpha} + C_{l1}\sum_{m=L+1}^{M} m^{-2\alpha} \to 0,$$

where the last inequality follows from the fact that $\sum_{s=1}^{m}|w_m| \le C_{l1} < \infty$ uniformly in N by Assumption 4. Then $\sum_{m=L+1}^{M} m^{-2\alpha} \to 0$ because $L \to \infty$ and $\sum_{m=1}^{M} m^{-2\alpha} < \infty$ uniformly in M. *Q.E.D.*

In what follows, $\sum_i$ and $\sum_{i\neq j}$ signify $\sum_{i=1}^{N}$ and $\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$, respectively.

LEMMA A.5: *Suppose that Assumptions* 1–3 *are satisfied. Then we have the following equalities*:

(i) $\operatorname{tr}(P(W)) = \sum_{m=1}^{M} w_m m = K'W$ (*Hansen* (2007, *Lemma* 1.1)).

(ii) $\sum_i (P_{ii}(W))^2 = o_p(K'W^+)$.

(iii) $\sum_{i\neq j} P_{ii}(W)P_{jj}(W) = (K'W)^2 + o_p(K'W^+)$.

(iv) $\sum_{i\neq j} P_{ij}(W)P_{ij}(W) = \sum_{m=1}^{M}\sum_{l=1}^{M} w_m w_l \min(l,m) + o_p(K'W) = W'\Gamma W + o_p(K'W)$.

(v) $\sum_{i\neq j} P_{ij}(W) = O_p(N - K'W)$.

(vi) $h = f'\epsilon/\sqrt{N} = O_p(1)$ *and* $H = f'f/N = O_p(1)$ (*Donald and Newey* (2001, *Lemma* A.2 (v))).

PROOF: We do not provide the proofs of parts (i) and (vi), as the proofs are available in Hansen (2007) and Donald and Newey (2001). For part (ii), first we note that $A_{ii} \le B_{ii}$ if $A \le B$, which implies that $P_{l,ii} \le P_{M,ii}$ for $l \le M$. Then Assumption 3 and Lemma A.5(i) imply

$$\sum_i (P_{ii}(W))^2 = \sum_{i=1}^{N}\sum_{m,l=1}^{M} w_m w_l P_{l,ii}P_{m,ii}$$

$$\le \sum_{i=1}^{N}\sum_{m,l=1}^{M} |w_m||w_l| P_{l,ii}P_{m,ii}$$

$$\begin{aligned}
&\leq \max_i(P_{M,ii})\left(\sum_{m=1}^{M}|w_l|\right)\sum_{i=1}^{N}\sum_{m=1}^{M}|w_m|P_{m,ii} \\
&\leq C\max_i(P_{M,ii})\operatorname{tr}P(W^+) \\
&= o_p(1)(K'W^+) = o_p(K'W^+),
\end{aligned}$$

where $\sum_{m=1}^{M}|w_l| \leq C_{l1}$ for some $C_{l1} < \infty$ was used and the bound holds uniformly for all N by Assumption 4. Also these results imply

$$\begin{aligned}
\sum_{i\neq j}P_{ii}(W)P_{jj}(W) &= \sum_i P_{ii}(W)\sum_j P_{jj}(W) - \sum_i (P_{ii}(W))^2 \\
&= (K'W)^2 + o_p(K'W^+),
\end{aligned}$$

which shows part (iii).

To show part (iv), first we observe that

$$\sum_{i\neq j}P_{ij}(W)P_{ij}(W) = \operatorname{tr}(P(W)P(W)) - \sum_i (P_{ii}(W))^2.$$

Now $\operatorname{tr}(P(W)P(W)) = \sum_{m=1}^{M}\sum_{l=1}^{M}w_m w_l \min(l,m)$ by Lemma 1.2 of Hansen (2007). Thus, by combining this result with part (ii) of this lemma, we get

$$\sum_{i\neq j}P_{ij}(W)P_{ij}(W) = \sum_{m=1}^{M}\sum_{l=1}^{M}w_m w_l \min(l,m) + o_p(K'W^+).$$

For part (v), note that

$$\sum_{i\neq j}P_{ij}(W) = \mathbf{1}_N' P(W)\mathbf{1}_N - \operatorname{tr}(P(W)),$$

where $\mathbf{1}_N' P_m \mathbf{1}_N \leq \mathbf{1}_N'\mathbf{1}_N = N$ by the fact that P_m is an idempotent matrix. Then note that

$$\begin{aligned}
\mathbf{1}_N' P(W)\mathbf{1}_N - \operatorname{tr}(P(W)) &= |\mathbf{1}_N' P(W)\mathbf{1}_N| - \operatorname{tr}(P(W)) \\
&\leq \sum_{m=1}^{M}|w_m||\mathbf{1}_N' P_m\mathbf{1}_N| - \operatorname{tr}(P(W)) \\
&\leq CN - K'W
\end{aligned}$$

such that $\sum_{i\neq j}P_{ij}(W) = O_p(N - K'W) = O_p(N)$. *Q.E.D.*

Let $e_f(W) = f'(I - P(W))(I - P(W))f/N$ and $\Delta(W) = \mathrm{tr}(e_f(W))$.

LEMMA A.6: *Suppose that Assumptions* 1–3, 4, *and* 5(i) *are satisfied. Then the following statements hold*:

(i) $\Delta(W) = o_p(1)$.

(ii) $f'(I - P(W))\epsilon/\sqrt{N} = O_p(\Delta(W)^{1/2})$.

(iii) $E[u'P(W)\epsilon|z] = \sigma_{u\epsilon}K'W$.

(iv) $E[u'P(W)\epsilon\epsilon'P(W)u|z] = \sigma_{u\epsilon}\sigma'_{u\epsilon}(K'W)^2 + (\sigma_\epsilon^2\Sigma_u + \sigma_{u\epsilon}\sigma'_{u\epsilon})(W'\Gamma W) + \mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u'_i]\sum_i(P_{ii}(W))^2$.

(v) $E[f'\epsilon\epsilon'P(W)u|z] = \sum_i f_iP_{ii}(W)E[\epsilon_i^2u'_i] = O_p(K'W^+)$.

(vi) *Let* $g(W): W \to \mathbb{R}$ *with* $g(W) > 0$ *be a function of* W *such that* $g(W) \to \infty$ *as* $N \to \infty$. *Then* $\sqrt{g(W)\Delta(W)}/\sqrt{N} = O_p(g(W)/N + \Delta(W))$.

(vii) $E[hh'H^{-1}u'f/N|z] = \sum_i f_if'_iH^{-1}E[\epsilon_i^2u_i]f'_i/N^2 = O_p(1/N)$ (*Donald and Newey* (2001, *Lemma* A.3 (vii))).

(viii) $E[f'(I - P(W))\epsilon\epsilon'P(W)u/N|z] = f'(I - P(W))\mu(W)/N = o_p((K'W^+)/N + \Delta(W))$.

(ix) $E[f'\epsilon\epsilon'fH^{-1}u'P(W)u|z]/N^2 = O_p(1/N) + \sigma_\epsilon^2\Sigma_uK'W/N$.

(x) $E[f'\epsilon\epsilon'P(W)uH^{-1}(u'f + f'u)|z]/N^2 = O_p(1/N) + (K'W/N)(\sum_i f_i \times \sigma'_{u\epsilon}H^{-1}\sigma_{u\epsilon}f_i/N + \sum_i f_i\sigma'_{u\epsilon}H^{-1}f_i\sigma'_{u\epsilon}/N)$.

(xi) $E[u'P(W)\epsilon\epsilon'fH^{-1}(u'f + f'u)|z]/N^2 = O_p(1/N) + (K'W/N) \times (d\sigma_{u\epsilon}\sigma'_{u\epsilon} + \sigma_{u\epsilon}\sum_i f'_iH^{-1}\sigma_{u\epsilon}f'_i/N)$.

(xii) $W'\Gamma W \le CK'W^+$.

PROOF: Let $\tilde{\gamma}_m = \mathrm{tr}(f'(I - P_m)f)/N$. By construction $\tilde{\gamma}_m \ge 0$. Write

$$\mathrm{tr}\big(f'(I - P(W))(I - P(W))f\big)/N = W'AW,$$

where

$$A = \begin{pmatrix} \tilde{\gamma}_1 & \tilde{\gamma}_2 & \cdots \\ \tilde{\gamma}_2 & \tilde{\gamma}_2 & \\ \vdots & & \ddots \end{pmatrix}.$$

It follows that

$$W'AW = \left(\sum_{m=1}^{M-1}\left(\sum_{s=1}^{m} w_s\right)^2(\tilde{\gamma}_m - \tilde{\gamma}_{m+1})\right) + \tilde{\gamma}_M \tag{A.9}$$
$$= \left(\sum_{m=1, m\notin \bar{J}}^{M-1}\left(\sum_{s=1}^{m} w_s\right)^2(\tilde{\gamma}_m - \tilde{\gamma}_{m+1})\right) + \tilde{\gamma}_M + o_p(1),$$

where the second equality holds by Assumption 2(iii) such that

$$W'AW \leq \sum_{m=1, m\notin \bar{J}}^{M-1} \left(\sum_{s=1}^{m} w_s \right)^2 \tilde{\gamma}_m + o_p(1) = \sum_{m=1, m\notin \bar{J}}^{M-1} \left(\sum_{s=1}^{m} w_s \right)^2 \frac{\tilde{\gamma}_m}{m^{-2\alpha}} m^{-2\alpha}$$
$$\leq \sup_{m\leq M} (m^{2\alpha} \tilde{\gamma}_m) \sum_{m=1, m\notin \bar{J}}^{M-1} \left(\sum_{s=1}^{m} w_s \right)^2 m^{-2\alpha},$$

where $\sup_{m\leq M}(m^{2\alpha}\tilde{\gamma}_m) = O_p(1)$ by Assumption 2(ii). For a sequence $L \leq M$, $L \to \infty$, and $L/N \leq M/N \to 0$ satisfying Assumption 4(ii) it follows that $\sum_{m=1, m\notin \bar{J}}^{M} (\sum_{s=1}^{m} w_s)^2 m^{-2\alpha} = o(1)$ by Lemma A.4. This implies that $\operatorname{tr}(f'(I - P(W))(I - P(W))f)/N = \Delta(W) = o_p(1)$.

Next, we observe that $E[f'(I - P(W))\epsilon/\sqrt{N}] = 0$ and

$$E\left[\frac{f'(I - P(W))\epsilon}{\sqrt{N}} \frac{\epsilon'(I - P(W))f}{\sqrt{N}} \Big| z \right]$$
$$= \sigma_\epsilon^2 \frac{f'(I - P(W))(I - P(W))f}{N} = \sigma_\epsilon^2 e_f(W).$$

Therefore, $f'(I - P(W))\epsilon/\sqrt{N} = O_p(\Delta(W)^{1/2})$ by the Chebyshev inequality. This shows part (ii).

For part (iii),

$$E[u'P(W)\epsilon|z] = \sum_{i=1}^{N} P_{ii}(W) E[u_i \varepsilon_i]$$
$$= \sigma_{u\epsilon} \operatorname{tr}(P(W)) = \sigma_{u\epsilon} K'W.$$

To give part (iv), observe that $E[u_i P_{ij}(W) \epsilon_j \epsilon_k P_{kl}(W) u_l'] = 0$ if one of (i, j, k, l) is different from all the rest. Also $E[\epsilon_i^2 u_i u_i']$ is bounded by Assumption 1. Therefore, we have

$$E[u'P(W)\epsilon\epsilon'P(W)u|z]$$
$$= \sum_i (P_{ii}(W))^2 E[\epsilon_i^2 u_i u_i'] + \sum_{i\neq j} E[u_i P_{ii}(W) \epsilon_i \epsilon_j P_{jj}(W) u_j' |z]$$
$$+ \sum_{i\neq j} E[u_i P_{ij}(W) \epsilon_j \epsilon_i P_{ij}(W) u_j' |z]$$
$$+ \sum_{i\neq j} E[u_i P_{ij}(W) \epsilon_j^2 P_{ji}(W) u_i' |z]$$

$$
\begin{aligned}
&= E[\epsilon_i^2 u_i u_i'] \sum_i (P_{ii}(W))^2 + \sigma_{u\epsilon}\sigma_{u\epsilon}' \sum_{i\neq j} P_{ii}(W)P_{jj}(W) \\
&\quad + (\sigma_\epsilon \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}') \sum_{i\neq j} P_{ij}(W)P_{ij}(W) \\
&= \mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] \sum_i (P_{ii}(W))^2 + \sigma_{u\epsilon}\sigma_{u\epsilon}'(K'W)^2 \\
&\quad + (\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')(W'\Gamma W)
\end{aligned}
$$

by Lemma A.5(iii) and (iv) and noting that $\mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] = E[\epsilon_i^2 u_i u_i'] - \sigma_\epsilon^2 \Sigma_u - 2\sigma_{u\epsilon}\sigma_{u\epsilon}'$.

Assumption 1 also implies

$$
E[f'\epsilon\epsilon' P(W)u|z] = \sum_{i,j,k} f_i P_{jk}(W) E[\epsilon_i \epsilon_j u_k'] = \sum_i f_i P_{ii}(W) E[\epsilon_i^2 u_i']
$$

and, furthermore, together with Assumption 3,

$$
\left\| \sum_i f_i P_{ii}(W) E[\epsilon_i^2 u_i'] \right\| \le \sum_i |P_{ii}(W)| \cdot \|f_i\| \cdot \|E[\epsilon_i^2 u_i']\| = O_p(K'W^+),
$$

which gives part (v).

To prove part (vi), first we consider the function of a: $g(W)/a + a$ for $a \in \mathbb{R}$, which is convex and the minimum value of which is $2\sqrt{g(W)}$ with the minimizer $a = \sqrt{g(W)}$. If $\Delta(W) = 0$, then $(\sqrt{\Delta(W)/N})/(g(W)/N + \Delta(W)) = 0$ and for $\Delta(W) \neq 0$,

$$
\text{(A.10)} \qquad \frac{\sqrt{\Delta(W)/N}}{g(W)/N + \Delta(W)} = \left(\frac{g(W)}{\sqrt{\Delta(W)N}} + \sqrt{\Delta(W)N} \right)^{-1} \le \frac{1}{2\sqrt{g(W)}} \to 0
$$

as $g(W) \to \infty$.

For part (viii), let $Q(W) = I - P(W)$ with (i, j)th element denoted by $Q_{ij}(W)$ and, for some a and b, let $f_{i,a} = f_a(z_i)$ and $\mu_{i,b}(W) = E[\epsilon_i^2 u_{ib}]P_{ii}(W)$. Now the (a, b)th element of $E[f'(I - P(W))\epsilon\epsilon' P(W)u|z]$ satisfies

$$
\begin{aligned}
&\left| E\left[\sum_{i,j,k,l} f_{i,a} Q_{ij}(W) \epsilon_j \epsilon_k P_{kl}(W) u_{lb} \middle| z \right] \right| \\
&\quad = \left| \sum_{i,j} f_{i,a} Q_{ij}(W) E[\epsilon_j^2 u_{jb}] P_{jj}(W) \right| \\
&\quad = |f_a' Q(W) \mu_b(W)| \le |f_a' Q(W) Q(W) f_a|^{1/2} |\mu_b'(W) \mu_b(W)|^{1/2},
\end{aligned}
$$

where the inequality is the Cauchy–Schwarz inequality. Now $|f_a'Q(W)Q(W)f_a|^{1/2} = O_p((N\Delta(W))^{1/2})$ by the definition of $\Delta(W)$. For some constant C, $|\mu_b'(W)\mu_b(W)| \leq C\sum_i (P_{ii}(W))^2$ by Assumption 1 and applying Lemma A.5(ii), we have $|\mu_b'(W)\mu_b(W)| = o_p(K'W^+)$. Therefore, we have

$$\begin{aligned}
&E\big[f'(I-P(W))\epsilon\epsilon'P(W)u/N|z\big] \\
&\quad = O_p\big((N\Delta(W))^{1/2}\big)o_p(\sqrt{K'W^+})O_p(1/N) \\
&\quad = o_p\big(\Delta(W)^{1/2}\sqrt{K'W^+}/\sqrt{N}\big) \\
&\quad = o_p((K'W^+)/N + \Delta(W)),
\end{aligned}$$

where the last equality follows from the fact that

$$\Delta(W)^{1/2}\sqrt{K'W^+}/\sqrt{N} \leq ((K'W^+)/N + \Delta(W))/2$$

by (A.10). In addition if we define $\mu_i(W) = E[\epsilon_i^2 u_i]P_{ii}(W)$ and $\mu(W) = (\mu_1(W)', \ldots, \mu_n(W)')'$, then

$$E\big[f'(I-P(W))\epsilon\epsilon'P(W)u/N|z\big] = f'(I-P(W))\mu(W)/N.$$

For part (ix), we have the decomposition

$$\begin{aligned}
&E[f'\epsilon\epsilon' f H^{-1}u'P(W)u|Z]/N^2 \\
&\quad = \sum_i f_i f_i' H^{-1}E[\epsilon_i^2 u_i u_i']P_{ii}(W)/N^2 \\
&\qquad + 2\sum_{i\neq j} f_i f_j' H^{-1}E[\epsilon_i u_i]E[\epsilon_j u_j']P_{ij}(W)/N^2 \\
&\qquad + \sum_{i\neq j} f_i f_i' H^{-1}E[\epsilon_i^2]E[u_j u_j']P_{jj}(W)/N^2.
\end{aligned}$$

The boundedness of $f_i f_i' H^{-1}P_{ii}(W)$ implies that

$$\sum_i f_i f_i' H^{-1}E[\epsilon_i^2 u_i u_i']P_{ii}(W)/N^2 = O_p(1/N).$$

Let $f_{a,i}$ be the ath element of f_i. Then we have

$$\begin{aligned}
\left|\sum_{i,j} f_{a,i}f_{a,j}P_{ij}(W)/N^2\right| &\leq \sum_{m=1} |w_m|(f_a'P_m f_a)/N^2 \\
&\leq \sum_{m=1} |w_m|(f_a'f_a)/N^2 = O_p(1/N).
\end{aligned}$$

This implies that

$$
\begin{aligned}
&\sum_{i\neq j} f_i f_j' H^{-1} E[\epsilon_i u_i] E[\epsilon_j u_j'] P_{ij}(W)/N^2 \\
&\quad = \sum_{i,j} f_i f_j' H^{-1} E[\epsilon_i u_i] E[\epsilon_j u_j'] P_{ij}(W)/N^2 \\
&\qquad - \sum_{i} f_i f_i' H^{-1} E[\epsilon_i u_i] E[\epsilon_i u_i'] P_{ii}(W)/N^2 \\
&\quad = O_p(1/N).
\end{aligned}
$$

Last, we have

$$
\begin{aligned}
&\sum_{i\neq j} f_i f_i' H^{-1} E[\epsilon_i^2] E[u_j u_j'] P_{jj}(W)/N^2 \\
&\quad = \left(\sum_i f_i f_i'\right) H^{-1} \sigma_\epsilon^2 \Sigma_u \left(\sum_j P_{jj}(W)\right) \Big/ N^2 \\
&\qquad - \sum_i f_i f_i' H^{-1} \sigma_\epsilon^2 \Sigma_u P_{ii}(W)/N^2 \\
&\quad = \sigma_\epsilon^2 \Sigma_u K'W/N + O_p(1/N).
\end{aligned}
$$

Therefore, we have

$$
E[f'\epsilon\epsilon' f H^{-1} u' P(W) u | Z]/N^2 = \sigma_\epsilon^2 \Sigma_u K'W/N + O_p(1/N).
$$

For part (x), using again Lemma A.5(v) as before,

$$
\begin{aligned}
&E[f'\epsilon\epsilon' P(W) u H^{-1} u' f | z]/N^2 \\
&\quad = \sum_i f_i P_{ii}(W) E[\epsilon_i^2 u_i' H^{-1} u_i | z] f_i'/N^2 \\
&\qquad + \sum_{i\neq j} f_i P_{jj}(W) E[\epsilon_j u_j'] H^{-1} E[u_i \epsilon_i] f_i'/N^2 \\
&\qquad + \sigma_\epsilon^2 \sum_{i\neq j} f_i P_{ij}(W) E[u_j' H^{-1} u_j | z] f_j'/N^2 \\
&\qquad + \sigma_\epsilon^2 \sum_{i\neq j} f_j P_{ji}(W) E[u_j' H^{-1} u_j] f_i'/N^2 \\
&\quad = O_p(1/N) + \sum_{i\neq j} f_i P_{jj}(W) E[\epsilon_j u_j'] H^{-1} E[u_i \epsilon_i] f_i'/N^2 \\
&\quad = O_p(1/N) + (K'W/N) \sum_i f_i \sigma_{u\epsilon}' H^{-1} \sigma_{u\epsilon} f_i/N
\end{aligned}
$$

and

$$
\begin{aligned}
&E[f'\epsilon\epsilon' P(W)uH^{-1}f'u|z]/N^2 \\
&= \sum_i f_i P_{ii}(W) E[\epsilon_i^2 u_i' H^{-1} f_i u_i'|z]/N^2 \\
&\quad + \sum_{i\neq j} f_i P_{jj}(W) E[\epsilon_j u_j'] H^{-1} f_i E[u_i'\epsilon_i]/N^2 \\
&\quad + \sigma_\epsilon^2 \sum_{i\neq j} f_i P_{ij}(W) E[u_j H^{-1} f_j u_j'|z]/N^2 \\
&\quad + \sigma_\epsilon^2 \sum_{i\neq j} f_j P_{ji}(W) E[u_j H^{-1} f_i u_j'|z]/N^2 \\
&= O_p(1/N) + \sum_{i\neq j} f_i P_{jj}(W) E[\epsilon_j u_j'] H^{-1} f_i E[u_i'\epsilon_i]/N^2 \\
&= O_p(1/N) + (K'W)/N \sum_i f_i \sigma_{u\epsilon}' H^{-1} f_i \sigma_{u\epsilon}'/N.
\end{aligned}
$$

For part (xi), with the same arguments, it holds that

$$
\begin{aligned}
&E[u'P(W)\epsilon\epsilon' f H^{-1} f'u|z]/N^2 \\
&= \sum_i P_{ii}(W) E[\epsilon_i^2 u_i f_i H^{-1} u_i f_i'|z]/N^2 \\
&\quad + \sum_{i\neq j} P_{jj}(W) E[\epsilon_j u_j] f_i' H^{-1} f_i E[u_i'\epsilon_i]/N^2 \\
&\quad + \sigma_\epsilon^2 \sum_{i\neq j} P_{ij}(W) E[u_j f_i' H^{-1} f_i u_i'|z]/N^2 \\
&\quad + \sum_{i\neq j} P_{ij}(W) E[u_j \varepsilon_j] f_j' H^{-1} f_i E[u_i'\varepsilon_i]/N^2 \\
&= O_p\left(\frac{1}{N}\right) + \frac{K'W}{N} \sigma_{u\epsilon}\sigma_{u\epsilon}' \frac{1}{N} \sum_{i=1}^n f_i' H^{-1} f_i \\
&= O_p\left(\frac{1}{N}\right) + \frac{K'W}{N} \sigma_{u\epsilon}\sigma_{u\epsilon}' \operatorname{tr}\left(H^{-1} \frac{1}{N} \sum_i f_i f_i'\right) \\
&= O_p\left(\frac{1}{N}\right) + d\frac{K'W}{N} \sigma_{u\epsilon}\sigma_{u\epsilon}'
\end{aligned}
$$

and arguments similar to before give

$$
\begin{aligned}
&E[u'P(W)\epsilon\epsilon' f H^{-1}u'f|z]/N^2 \\
&\quad = O_p(1/N) + \sum_{i\neq j} P_{jj}(W)E[\epsilon_j u_j]f_i' H^{-1}E[u_i\epsilon_i]f_i'/N^2 \\
&\quad = O_p\left(\frac{1}{N}\right) + \frac{K'W}{N}\sigma_{u\epsilon}\frac{1}{N}\sum_i f_i' H^{-1}\sigma_{u\epsilon} f_i'.
\end{aligned}
$$

For part (xii), note that

$$
\begin{aligned}
W'\Gamma W &= \sum_{m=1}^{M}\left(\sum_{j=m}^{M} w_j\right)^2 \\
&\leq \sum_{m=1}^{M}\sum_{j=m}^{M}|w_j|\left|\sum_{j=m}^{M} w_j\right| \\
&\leq C\sum_{m=1}^{M}|w_m|m = CK'W^+,
\end{aligned}
$$

where the second inequality follows from the condition $\sup_{k\leq M}|\sum_{m=k}^{M} w_m| \leq C_{l1} < \infty$, which holds uniformly in M. *Q.E.D.*

LEMMA A.7: *Assume that Assumptions* 1, 2, 3, *and* 4 *hold. Let*

$$\Xi(W) = \operatorname{tr}\big(f'(I-P(W))f/N\big). \tag{A.11}$$

Let $\rho_{W,N} = \operatorname{tr}(S(W))$, *where* $S(W)$ *is defined in* (A.6). *Then we have*

$$(\Xi(W))^2 = o_p(\rho_{W,N}).$$

We note that the result holds when $S(W)$ *is defined in* (A.2).

REMARK A.3: Considering the set $\bar{J}$ in Assumption 2 is important because the optimal weighting vector has a structure such that w_j does not converge to 0 if $f'(P_m - P_{m+1})f/N = 0$. Thus, the optimal weighting vector does not satisfy $\sup_{j\leq L}|\sum_{s=1}^{j} w_s| = O(1/\sqrt{N})$ in general.

PROOF OF LEMMA A.7: Let $\tilde{\gamma}_m = \operatorname{tr}(f'(I-P_m)f/N)$ and let A be the $M\times M$ matrix whose (i, j)th element is $\min(\tilde{\gamma}_i, \tilde{\gamma}_j) = \tilde{\gamma}_{\max(i,j)}$. Let e_1 be the first unit vector. We write

$$\Xi(W) = W'Ae_1, \quad \Delta(W) = W'AW.$$

Let $W_1 = (w_1, \dots, w_L, 0 \dots, 0)$ and $W_2 = (0, \dots, 0, w_{L+1}, \dots, w_M)$. We have the decomposition

$$
\begin{aligned}
(\Xi(W))^2 &= W_1' A e_1 e_1' A W_1 + 2W_1' A e_1 e_1' A W_2 + W_2' A e_1 e_1' A W_2, \\
\Delta(W) &= W_1' A W_1 + 2W_1' A W_2 + W_2' A W_2.
\end{aligned}
$$

First, we consider

$$
\begin{aligned}
W_1' A W_1 &= \sum_{j=1}^{L-1} \left(\sum_{s=1}^{j} w_s \right)^2 (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) + \left(\sum_{s=1}^{L} w_s \right)^2 \tilde{\gamma}_L \\
&= \sum_{j \notin \bar{J}, j \le L} \left(\sum_{s=1}^{j} w_s \right)^2 (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) + \left(\sum_{s=1}^{L} w_s \right)^2 \tilde{\gamma}_L \quad \text{wpa1} \\
&\le \sup_{j \notin \bar{J}, j \le L} \left(\sum_{s=1}^{j} w_s \right)^2 \left(\sum_{j \notin \bar{J}, j \le L-1} (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) + \tilde{\gamma}_L \right) \quad \text{wpa1} \\
&= \sup_{j \notin \bar{J}, j \le L} \left(\sum_{s=1}^{j} w_s \right)^2 \tilde{\gamma}_1 \quad \text{wpa1} \\
&= O_p(1/N).
\end{aligned}
$$

By Lemma A.3, $W' \Gamma W \to \infty$ so that

$$
W_1' A W_1 = O_p(1/N) = o(W' \Gamma W / N) = o(\rho_{W,N}).
$$

Since $|W_1' A W_2| \le (W_1' A W_1)^{1/2} (W_2' A W_2)^{1/2}$ by the Cauchy–Schwarz inequality, we have $\Delta(W) = W_2' A W_2 + o_p(\rho_{W,N})$. Next, we consider

$$
\begin{aligned}
W_1' A e_1 e_1' A W_1 &= (W_1' A e_1)^2 \\
&\le (W_1' A W_1)(e_1' A e_1) \\
&= (W_1' A W_1) \tilde{\gamma}_1 = O_p(W_1' A W_1) = o_p(\rho_{W,N}),
\end{aligned}
$$

where the inequality is that of Cauchy–Schwarz. We examine the order of $W_2' A e_1 e_1' A W_2$. We observe that

$$
W_2' A e_1 = \sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right) (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) + \left(\sum_{s=L+1}^{M} w_s \right) \tilde{\gamma}_M
$$

and

$$W_2' A W_2 = \sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right)^2 (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) + \left(\sum_{s=L+1}^{M} w_s \right)^2 \tilde{\gamma}_M .$$

These formulas imply that

$$\begin{aligned} & W_2' A e_1 - W_2' A W_2 \\ & \quad = \sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right) \left(1 - \left(\sum_{s=L+1}^{j} w_s \right) \right) (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) \\ & \qquad + \left(\sum_{s=L+1}^{M} w_s \right) \left(1 - \left(\sum_{s=L+1}^{M} w_s \right) \right) \tilde{\gamma}_M \\ & \quad = \sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right) \left(1 - \left(\sum_{s=L+1}^{j} w_s \right) \right) (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) \\ & \qquad + o_p(\rho_{W,N}), \end{aligned}$$

where

$$\left| \left(\sum_{s=L+1}^{M} w_s \right) \left(1 - \left(\sum_{s=L+1}^{M} w_s \right) \right) \right| \tilde{\gamma}_M \leq C \tilde{\gamma}_M = o_p(\rho_{W,N}).$$

We observe that, by the Cauchy–Schwarz inequality,

$$\begin{aligned} & \left(\sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right) \left(1 - \left(\sum_{s=L+1}^{j} w_s \right) \right) (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) \right)^2 \\ & \quad \leq \left(\sum_{j=L+1}^{M} \left(\sum_{s=L+1}^{j} w_s \right)^2 (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) \right) \\ & \qquad \times \left(\sum_{j=L+1}^{M} \left(1 - \left(\sum_{s=L+1}^{j} w_s \right) \right)^2 (\tilde{\gamma}_j - \tilde{\gamma}_{j+1}) \right) \\ & \quad \leq W_2' A W_2 \cdot C(\tilde{\gamma}_L - \tilde{\gamma}_M) + o_p(\rho_{W,N}) = o_p(\rho_{W,N}) \end{aligned}$$

since $W_2 A W_2 = O(\rho_{W,N})$ and $\tilde{\gamma}_L - \tilde{\gamma}_M = o_p(1)$. It therefore follows that

$$(W_2' A e_1 - W_2' A W_2)^2 = o_p(\rho_{W,N}).$$

Therefore,

$$W_2'Ae_1e_1A'W_2 = (W_2'AW_2 + W_2'Ae_1 - W_2'AW_2)^2$$
$$\le 2(W_2AW_2)^2 + 2(W_2'Ae_1 - W_2'AW_2)^2 = o_p(\rho_{W,N}).$$

Last, by the Cauchy–Schwarz inequality, we have

$$W_1'Ae_1e_1'AW_2 = o_p(\rho_{W,N}).$$

To sum up, we have

$$(\Xi(W))^2 = W_1'Ae_1e_1'AW_1 + 2W_1'Ae_1e_1'AW_2 + W_2'Ae_1e_1'AW_2$$
$$= o_p(\rho_{W,N}). \qquad Q.E.D.$$

LEMMA A.8: *If Assumptions* 1–8 *hold and* $\Omega = \Omega_U = \{W \in l_1 | W'\mathbf{1}_M = 1\}$, *where* M *satisfies the constraints in Assumption* 10 *and* $W = (w_1, \ldots, w_M)$, *it follows that*

$$\inf_{W\in\Omega} S_\lambda(W) = O_p\big(N^{-2\alpha/(2\alpha+1)}\big),$$

where $S_\lambda(W) = \lambda' S(W)\lambda$ *and* $S(W)$ *is defined in* (A.2).

PROOF: Consider a sequence $\tilde{W}$, where $w_M = 2$, $w_{2M} = -1$, and $w_j = 0$ for $j \neq M, 2M$ and $M = \lfloor N^{1/(2\alpha+1)} \rfloor$. Clearly, $\mathbf{1}'\tilde{W} = 1$ and $\tilde{W} \in l_1$ for all N such that $\tilde{W} \in \Omega$. We note that $K'\tilde{W} = 0$. It follows that

$$S_\lambda(\tilde{W}) = \lambda' H^{-1}\bigg(b_\sigma \frac{(\tilde{W}'\Gamma\tilde{W})}{N}$$
$$+ \sigma_\epsilon^2 \frac{f'(I - P(\tilde{W}))(I - P(\tilde{W}))f}{N}\bigg)H^{-1}\lambda,$$

where

$$\frac{(\tilde{W}'\Gamma\tilde{W})}{N} = \frac{2M}{N} = O\big(N^{-2\alpha/(2\alpha+1)}\big)$$

and

$$\frac{\mathrm{tr}(f'(I - P(\tilde{W}))(I - P(\tilde{W}))f)}{N} = 4\tilde{\gamma}_M - 3\tilde{\gamma}_{2M}$$
$$= O_p(M^{-2\alpha}) = O_p\big(N^{-2\alpha/(2\alpha+1)}\big),$$

where $\tilde{\gamma}_m = \mathrm{tr}(f'(I - P_m)f/N)$. This argument shows that $\inf_{W\in\Omega} S_\lambda(W) \le CN^{-2\alpha/(2\alpha+1)}$.

To show that the rate is sharp, suppose that there is an $\varepsilon > 0$ such that

$$\inf_{W\in\Omega} S_\lambda(W) = O_p\big(N^{(-2\alpha(1+\varepsilon))/(2\alpha+1)}\big).$$

Take any W such that, for $M = \lfloor N^{(1+\delta)/(2\alpha+1)} \rfloor$, where $0 < \delta < \varepsilon/2$,

$$\begin{aligned}\text{(A.12)}\quad \operatorname{tr}\left(\frac{f'(I-P(\tilde W))(I-P(\tilde W))f}{N}\right) &= \sum_{j=1}^{M}\left(\sum_{i=1}^{j} w_i\right)^2(\tilde\gamma_j - \tilde\gamma_{j+1}) + \tilde\gamma_M \\ &= O_p\big(N^{(-2\alpha(1+\varepsilon))/(2\alpha+1)}\big),\end{aligned}$$

where we use formula (A.9). Let J_M be the set of integers j such that $1 \le j \le M$ for which $j^{2\alpha+1}(\tilde\gamma_j - \tilde\gamma_{j+1}) > 0$. By the assumptions of the lemma, wpa1, $\sharp J_M = O(M)$ as $M \to \infty$, where $\sharp J_M$ is the cardinality of J_M. It follows that

$$\begin{aligned}\sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)^2(\tilde\gamma_j - \tilde\gamma_{j+1}) &\ge \sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)^2 M^{-(2\alpha+1)} \\ &\ge O\big(N^{(-(2\alpha+1)(1+\delta))/(2\alpha+1)}\big)\sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)^2,\end{aligned}$$

which together with (A.12) implies that

$$\sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)^2 = O\big(N^{(-2\alpha(\varepsilon-\delta)+1+\delta)/(2\alpha+1)}\big) = o(M).$$

Now, since

$$\begin{aligned}\text{(A.13)}\quad O(M) &= \sum_{j\in J_M} 1^2 \\ &= \sum_{j\in J_M}\left(\left(\sum_{i=1}^{j} w_i\right)^2 + 2\left(\sum_{i=1}^{j} w_i\right)\left(\sum_{i=j+1}^{M} w_i\right) + \left(\sum_{i=j+1}^{M} w_i\right)^2\right)\end{aligned}$$

and by the Cauchy–Schwarz inequality,

$$\begin{aligned}\left|\sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)\left(\sum_{i=j+1}^{M} w_i\right)\right| &\le \left(\sum_{j\in J_M}\left(\sum_{i=1}^{j} w_i\right)^2\right)^{1/2}\left(\sum_{j\in J_M}\left(\sum_{i=j+1}^{M} w_i\right)^2\right)^{1/2} \\ &= o(\sqrt{M})\left(\sum_{j\in J_M}\left(\sum_{i=j+1}^{M} w_i\right)\right)^{1/2},\end{aligned}$$

it follows that (A.13) can only hold if $\liminf_N \sum_{j\in J_M} (\sum_{i=j+1}^{M} w_i)^2/M > 0$. Then, for some $\eta > 0$ and N large enough, it follows that

$$W'\Gamma W = \sum_{j=0}^{M} \left(\sum_{m=j+1}^{M} w_m \right)^2 \geq M\eta = O\big(N^{(1+\delta)/(2\alpha+1)}\big)$$

such that $W'\Gamma W/N = O(N^{(-2\alpha+\delta)/(2\alpha+1)})$, which implies that $S_\lambda(W) = O(N^{(-2\alpha+\delta)/(2\alpha+1)})$, a contradiction to the assumption that $\inf_{W\in\Omega} S_\lambda(W) = O_p(N^{(-2\alpha(1+\varepsilon))/(2\alpha+1)})$. This argument establishes that $\inf_{W\in\Omega} S_\lambda(W) = O_p(N^{(-2\alpha)/(2\alpha+1)})$ is a sharp bound. *Q.E.D.*

LEMMA A.9: *Let*

$$\tilde{S}_\lambda(W) = \lambda' \hat{H}^{-1}\bigg(\hat{a}_\sigma \frac{(K'W)^2}{N} + \hat{b}_\sigma \frac{(W'\Gamma W)}{N} - \frac{K'W}{N}\hat{B}_N + \hat{\sigma}_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}\bigg)\hat{H}^{-1}\lambda.$$

If Assumptions 1–9 *hold, then for Ω as defined in Lemma* A.8, *it follows that*

$$\sup_{W\in\Omega} \frac{\tilde{S}_\lambda(W)}{S_\lambda(W)} - 1 = o_p(1),$$

where $S_\lambda(W) = \lambda' S(W)\lambda$ and $S(W)$ is defined in (A.2).

PROOF: We define the subset $\Omega_2 = \{W \in l_1 | -\infty < \liminf_N K'W \leq \limsup_N K'W < \infty\}$. Note that

$$\sup_{W\in\Omega\cap\Omega_2} \frac{K'W/N}{S_\lambda(W)} \to 0 \quad \text{and} \quad \sup_{W\in\Omega\cap\Omega_2} \frac{(K'W)^2/N}{S_\lambda(W)} \to 0 \tag{A.14}$$

by Lemma A.8 and the fact that $\{W \in l_1 | K'W = 0\} \in \Omega_2$. It now follows immediately that

$$\lambda'(\hat{H}^{-1}\hat{a}_\sigma \hat{H}^{-1} - H^{-1}a_\sigma H^{-1})\lambda \sup_{W\in\Omega\cap\Omega_2} \frac{(K'W)^2/N}{S_\lambda(W)} = o_p(1)$$

with the same argument holding for the term $\hat{B}_N K'W/N$. Define

$$S_{\lambda,\Omega_2}(W) = \lambda' H^{-1}\bigg(b_\sigma \frac{(W'\Gamma W)}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}\bigg) H^{-1}\lambda$$

and note that $S_{\lambda,\Omega_2}(W) \geq \lambda' H^{-1} b_\sigma H^{-1} \lambda (W'\Gamma W)/N$ as well as $S_{\lambda,\Omega_2}(W_N) \geq \sigma_\epsilon^2 \lambda' H^{-1} f'(I-P(W))(I-P(W)) f H^{-1}\lambda/N$. Thus, we have

$$\begin{aligned}
&\sup_{W\in\Omega\cap\Omega_2} \frac{(W'\Gamma W)/N}{S_\lambda(W)} \\
&\quad \leq \sup_{W\in\Omega\cap\Omega_2} \frac{(W'\Gamma W)/N}{S_{\lambda,\Omega_2}(W)} \sup_{W\in\Omega\cap\Omega_2} \frac{S_{\lambda,\Omega_2}(W)}{S_\lambda(W)} \\
&\quad \leq \frac{1}{\lambda' H^{-1} b_\sigma H^{-1}\lambda} \sup_{W\in\Omega\cap\Omega_2} \frac{S_{\lambda,\Omega_2}(W)}{S_\lambda(W)},
\end{aligned}$$

where $\sup_{W\in\Omega\cap\Omega_2} S_{\lambda,\Omega_2}(W_N)/S_\lambda(W_N) \to 1$ by (A.14). This implies that

$$\lambda'(\hat{H}^{-1}\hat{b}_\sigma \hat{H}^{-1} - H^{-1} b_\sigma H^{-1})\lambda \sup_{W\in\Omega\cap\Omega_2} \frac{(W'\Gamma W)/N}{S_\lambda(W)} = o_p(1).$$

Now consider

$$\begin{aligned}
&\lambda'(\hat{H}^{-1}\hat{\sigma}_\epsilon^2 - H^{-1}\sigma_\epsilon^2)\frac{f'(I-P(W))(I-P(W))f}{N}\hat{H}^{-1}\lambda \\
&\quad + \lambda' H^{-1}\sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}(\hat{H}^{-1} - H^{-1})\lambda,
\end{aligned}$$

where

$$\begin{aligned}
&\sup_{W\in\Omega\cap\Omega_2} \frac{|\lambda'(\hat{H}^{-1}\hat{\sigma}_\epsilon^2 - H^{-1}\sigma_\epsilon^2) f'(I-P(W))(I-P(W)) f\hat{H}^{-1}\lambda/N|}{S_\lambda(W)} \\
&\quad \leq \|\hat{H}^{-1}\lambda\| |\lambda'(\hat{H}^{-1}\hat{\sigma}_\epsilon^2 - H^{-1}\sigma_\epsilon^2)| \sup_{W\in\Omega} \frac{\|(I-P(W))f/\sqrt{N}\|^2}{\|(I-P(W))fH^{-1}\lambda/\sqrt{N}\|^2} \\
&\quad = o_p(1),
\end{aligned}$$

where

$$\sup_{W\in\Omega} \frac{\|(I-P(W))f/\sqrt{N}\|^2}{\|(I-P(W))fH^{-1}\lambda/\sqrt{N}\|^2} = O_p(1)$$

by Assumption 2. Together, these arguments show that

$$\sup_{W\in\Omega\cap\Omega_2} \frac{\tilde{S}_\lambda(W)}{S_\lambda(W)} - 1 = o_p(1).$$

For $W \in \Omega \cap \Omega_2^C$, where $\Omega_2^C = \{W \in l_1 | \liminf_N |K'W| = \infty\}$, it follows that

$$\sup_{W \in \Omega \cap \Omega_2^C} \frac{|K'W|/N}{(K'W)^2/N} \to 0$$

such that for

$$S_{\lambda,\Omega_2^C}(W_N) = \lambda' H^{-1}\left[a_\sigma \frac{(K'W)^2}{N} + b_\sigma \frac{(W'\Gamma W)}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}\right] H^{-1}\lambda,$$

it follows that

$$\sup_{W \in \Omega \cap \Omega_2^C} \frac{S_{\lambda,\Omega_2^C}(W)}{S_\lambda(W)} \to 1 \quad \text{as} \quad N \to \infty.$$

Then similar arguments as before can be used to show that

$$\sup_{W \in \Omega \cap \Omega_2} \frac{\tilde{S}_\lambda(W)}{S_\lambda(W)} - 1 = o_p(1).$$

Since $(\Omega_2 \cup \Omega_2^C) \cap \Omega = \Omega$, this establishes the claimed result. *Q.E.D.*

LEMMA A.10: *Let Assumptions* 1–10 *hold. Then it follows that*

$$\sup_{W \in \Omega} \frac{\hat{S}_\lambda(W)}{S_\lambda(W)} - 1 \to_p 0,$$

where $S_\lambda(W) = \lambda' S(W)\lambda$ *and* $S(W)$ *is defined in* (A.2).

PROOF: Without loss of generality assume that f_i is a scalar and $\lambda' H^{-1} = 1$ so that $\sigma_\lambda^2 = \sigma_u^2$. First consider

$$\begin{aligned} &\left\| (I - P(W))f/\sqrt{N} \right\|^2 - f'(I - P_M)f/N \\ &\quad = \left\| (P_M - P(W))f/\sqrt{N} \right\|^2 \end{aligned}$$

and note that

$$f'(I - P_M)f/N = O_p(M^{-2\alpha})$$

by Assumption 2. Together with Lemma A.8, this implies that

$$\sup_{W\in\Omega}\frac{\|(P_M-P(W))f/\sqrt{N}\|^2-\|(I-P(W))f/\sqrt{N}\|^2}{S_\lambda(W)}$$
$$\le\frac{\sup_{W\in\Omega}f'(I-P_M)f/N}{\inf_{W\in\Omega}S_\lambda(W)}$$
$$=O_p\big(M^{-2\alpha}N^{2\alpha/(2\alpha+1)}\big)=O_p\big(N^{-2\alpha\delta/(2\alpha+1)}\big)=o_p(1).$$

Combining these results with Lemma A.9, it is then sufficient to show that

$$\sup_{W\in\Omega}\big|\big\|(P_M-P(W))X/\sqrt{N}\big\|^2-\big\|(P_M-P(W))f/\sqrt{N}\big\|^2$$
$$-\sigma_u^2(M-2K'W+W'\Gamma W)/N\big|/S_\lambda(W)=o_p(1).$$

We note that in this expression we replace $\hat\sigma_u^2$ by σ_u^2, which is justified by the same arguments as in the proof of Lemma A.9 as long as $\hat\sigma_u^2-\sigma_u^2=o_p(N^{-\delta/(2\alpha+1)})$ because, under the assumptions of the lemma, it then follows that $(\hat\sigma_u^2-\sigma_u^2)M/N=o_p(N^{-2\alpha/(2\alpha+1)})=o_p(\inf_{W\in\Omega}S_\lambda(W))$ and the remaining terms involving σ_u^2 can be handled in the same way as in the proof of Lemma A.9. Now note that

$$\big\|(P_M-P(W))X/\sqrt{N}\big\|^2-\big\|(P_M-P(W))f/\sqrt{N}\big\|^2$$
$$=\big\|(P_M-P(W))u/\sqrt{N}\big\|^2+2u'(P_M-P(W))(P_M-P(W))f/N.$$

It follows that

$$E\big[u'(P_M-P(W))(P_M-P(W))u/N|z\big]$$
$$=\sigma_u^2\big(\mathrm{tr}(P_M)-2\,\mathrm{tr}(P(W))+\mathrm{tr}(P(W)P(W))\big)/N$$
$$=\sigma_u^2(M-2K'W+W'\Gamma W)/N$$

and

$$E\big[u'(P_M-P(W))(P_M-P(W))f/N|z\big]=0.$$

Moreover, we have the bound

$$\big|\big\|(P_M-P(W))u\big\|^2-\sigma_u^2(M-2K'W+W'\Gamma W)\big|$$
$$\le|u'P_Mu-\sigma_u^2M|$$
$$+\sup_{j\le M}|u'P_ju-\sigma_u^2j|\left(2\sum_{j=1}^M|w_j|+\sum_{j=1}^M\sum_{l=1}^M|w_j||w_l|\right),$$

where $\sum_{j=1}^{M}|w_j| \le C_{l1}$ uniformly in M is used. It follows for some $\vartheta > 1$ from Whittle (1960, Theorem 2) that for some constant C,

$$E[|u'P_j u - \sigma_u^2 j|^{2\vartheta}|z] \le CE[|u_i|^{2\vartheta}]^2(\mathrm{tr}(P_j P_j'))^{\vartheta} = CE[|u_i|^{2\vartheta}]^2 j^{\vartheta}$$

and thus for any $\eta > 0$ and some constant C, not necessarily the same as above,

$$\begin{aligned}
&\Pr\left[\frac{\sup_{W\in\Omega}|\|(P_M - P(W))u\|^2 - \sigma_u^2(M - 2K'W + W'\Gamma W)|/N}{\inf_{W\in\Omega} S_\lambda(W)} > \eta\right] \\
&\le C\frac{E[|u'P_M u - \sigma_u^2 M|^{2\vartheta}|z]}{\eta^{\vartheta}N^{2\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}} + 3C\sum_{j=1}^{M}\frac{E[|u'P_j u - \sigma_u^2 j|^{2\vartheta}|z]}{\eta^{\vartheta}N^{2\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}} \\
&\le C\frac{E[|u_i|^{2\vartheta}]^2(M^{\vartheta} + M^{\vartheta+1})}{\eta^{\vartheta}N^{2\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}} \\
&= O\left(N^{(1+\delta-\vartheta(1-\delta))/2\alpha+1}\right) = o(1).
\end{aligned}$$

Next, consider

$$\left|u'(P_M - P(W))(I - P(W))f/N\right| = \left|\sum_{i,j=1}^{M} w_i w_j u'\left(P_M - P_{\max(i,j)}\right)f/N\right|,$$

where

$$\left|\sum_{i,j=1}^{M} w_i w_j u'\left(P_M - P_{\max(i,j)}\right)f/N\right| \le \sum_{i=1}^{M-1}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1} - P_i)f/N|.$$

Let $K_n = N^{\lfloor(1-\varepsilon)/(2\alpha+1)\rfloor}$. Then

$$\begin{aligned}
\text{(A.15)}\quad &\sup_{W\in\Omega}\frac{\sum_{i=1}^{M-1}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1} - P_i)f/N|}{S_\lambda(W)} \\
&= \sup_{W\in\Omega}\frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1} - P_i)f/N|}{S_\lambda(W)} + o_p(1)
\end{aligned}$$

because

$$
\begin{aligned}
&\Pr\left(\sup_{W\in\Omega}\frac{\sum_{i=K_n+1}^{M-1}\left(\sum_{j=1}^{i}w_j\right)^2|u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)}>\eta\,\middle|\,z\right)\\
&\quad\le \Pr\left(\frac{\sup_{W\in\Omega}\sum_{i=K_n+1}^{M-1}\left(\sum_{j=1}^{i}w_j\right)^2|u'(P_{i+1}-P_i)f/N|}{\inf_{W\in\Omega}S_\lambda(W)}>\eta\,\middle|\,z\right)\\
&\quad\le \frac{CE[|u_i|^{2\vartheta}]\sum_{j=K_n+1}^{M}(f'(P_{j+1}-P_j)f/N)^{\vartheta}}{\eta^{\vartheta}N^{\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}},
\end{aligned}
$$

where the inequality follows from Markov's inequality, Lemma A.8, the fact that $|\sum_{j=1}^{i}w_j|$ is uniformly bounded on Ω, and Theorem 1 of Whittle (1960), which implies that

$$
E\big[|u'(P_{i+1}-P_i)f/N|^{2\vartheta}|z\big]\le CE[|u_i|^{2\vartheta}]N^{-\vartheta}(f'(P_{i+1}-P_i)f/N)^{\vartheta}. \tag{A.16}
$$

Now note that

$$
\begin{aligned}
&\frac{CE[|u_i|^{2\vartheta}]\sum_{j=K_n+1}^{M}(f'(P_{j+1}-P_j)f/N)^{\vartheta}}{\eta^{\vartheta}N^{\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}}\\
&\quad\le \frac{CE[|u_i|^{2\vartheta}](f'(I-P_{K_N})f/N)^{\vartheta}M}{\eta^{\vartheta}N^{\vartheta}N^{-4\alpha\vartheta/(2\alpha+1)}}\\
&\quad=O_p\big(K_n^{-2\alpha\vartheta}M/N^{\vartheta}N^{4\alpha\vartheta/(2\alpha+1)}\big)\\
&\quad=O_p\big(N^{-(2(1-\varepsilon)\alpha\vartheta)/(2\alpha+1)-\vartheta+(1+\delta)/(2\alpha+1)+4\alpha\vartheta/(2\alpha+1)}\big)\\
&\quad=o_p(1),
\end{aligned}
$$

which establishes (A.15). We thus turn to the lead term on the right hand side of (A.15). By the Cauchy–Schwarz inequality, we have

$$
|u'(P_{i+1}-P_i)f/N|\le (f'(P_{i+1}-P_i)f/N)^{1/2}(u'(P_{i+1}-P_i)u/N)^{1/2}.
$$

It now follows that

$$
\begin{aligned}
\text{(A.17)} \quad & \sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N| \\
& \leq \left(\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^4 f'(P_{i+1}-P_i)f/N\right)^{1/2} \\
& \quad \times \left(\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^4 u'(P_{i+1}-P_i)u/N\right)^{1/2} \\
& \leq \sup_{i\leq M}\left(\sum_{j=1}^{i} w_j\right)^2 \left(\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 f'(P_{i+1}-P_i)f/N\right)^{1/2} \\
& \quad \times \left(\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 u'(P_{i+1}-P_i)u/N\right)^{1/2},
\end{aligned}
$$

where $\sup_{i\leq M}(\sum_{j=1}^{i} w_j)^2 \leq C_{l1}^2 < \infty$ uniformly in M such that

$$
\begin{aligned}
\text{(A.18)} \quad & \left(\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 u'(P_{i+1}-P_i)u/N\right)^{1/2} \\
& \leq \sup_{W}\left(\sum_{j=1}^{i} |w_j|\right)^2 \left(\sum_{i=1}^{K_n} u'(P_{i+1}-P_i)u/N\right)^{1/2} \\
& \leq C\big(u'\big(P_{K_n+1}-P_1\big)u/N\big)^{1/2},
\end{aligned}
$$

where $W \in l_1$ was used to bound $\sup_W(\sum_{j=1}^{i}|w_j|)^2$. Let $\Omega_N \subset \Omega$ be the sequence of subsets of sequences in Ω for which $w_i = 0$ for all $i > N$. Clearly,

$$
\begin{aligned}
\text{(A.19)} \quad & \sup_{W\in\Omega} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)} \\
& = \sup_{W\in\Omega_N} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)}.
\end{aligned}
$$

Now fix an arbitrary $\omega > 0$ and define the sequence of sets

$$\Omega_{1,N} = \left\{ W \in \Omega_N \middle| \frac{\sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 f'(P_{i+1} - P_i) f/N}{N^{(-2\alpha+\varepsilon/2)/(2\alpha+1)}} \leq \omega \right\}$$

and let $\Omega_{1,N}^C$ be the complement of $\Omega_{1,N}$ in Ω_N such that $\Omega_N = (\Omega_N \cap \Omega_{1,N}) \cup (\Omega_N \cap \Omega_{1,N}^c)$. We note that $\Omega_{1,N}$ depends on the realizations for the instruments z.

As was demonstrated in the proof of Lemma A.9, as N tends to infinity, $S_\lambda(W) \geq \sigma_\epsilon^2 \lambda' H^{-1} f'(I - P(W))(I - P(W)) f H^{-1} \lambda / N$. Also note that

$$f'(I - P(W))(I - P(W)) f/N \geq \sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 f'(P_{i+1} - P_i) f/N.$$

Therefore, for N sufficiently large,

$$\begin{aligned}
&\sup_{W \in \Omega_N \cap \Omega_{1,N}^C} \frac{\sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 |u'(P_{i+1} - P_i) f/N|}{S_\lambda(W)} \\
&\leq \sup_{W \in \Omega_N \cap \Omega_{1,N}^C} \frac{\sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 |u'(P_{i+1} - P_i) f/N|}{\sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 f'(P_{i+1} - P_i) f/N} \\
&\leq \frac{C(u'(P_{K_n+1} - P_1) u/N)^{1/2}}{\inf_{W \in \Omega_N \cap \Omega_{1,N}^C} \left(\sum_{i=1}^{K_n} \left(\sum_{j=1}^{i} w_j \right)^2 f'(P_{i+1} - P_i) f/N \right)^{1/2}},
\end{aligned}$$

where

$$\inf_{W \in \Omega_N \cap \Omega_{1,N}^C} \frac{\left(\sum_{i \in J_{K_n}} \left(\sum_{j=1}^{i} w_j \right)^2 f'(P_{i+1} - P_i) f/N \right)^{1/2}}{N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}} \geq \sqrt{\omega}$$

by the construction of $\Omega_{1,N}$. It then follows that

$$\sup_{W\in\Omega_N\cap\Omega_{1,N}^C} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)} \leq \frac{C(u'(P_{K_n+1}-P_1)u/N)^{1/2}}{\sqrt{\omega}N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}}. \tag{A.20}$$

Second,

$$\sup_{W\in\Omega_N\cap\Omega_{1,N}} \sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 f'(P_{i+1}-P_i)f/N \leq \omega N^{(-2\alpha+\varepsilon/2)/(2\alpha+1)} \tag{A.21}$$

by the definition of $\Omega_{1,N}$ such that

$$\begin{aligned}\sup_{W\in\Omega_N\cap\Omega_{1,N}} &\frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)} \\ &\leq \frac{\sup_{W\in\Omega_N\cap\Omega_{1,N}} \sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{\inf_{W\in\Omega} S_\lambda(W)} \\ &\leq \sqrt{\omega}N^{(-\alpha+\varepsilon/4)/(2\alpha+1)} \frac{C(u'(P_{K_n+1}-P_1)u/N)^{1/2}}{\inf_{W\in\Omega} S_\lambda(W)}.\end{aligned} \tag{A.22}$$

It now follows for any random function $g_N(W)$ that

$$\begin{aligned}\sup_{W\in\Omega_N} g_N(W) &= \max\left(\sup_{W\in\Omega_N\cap\Omega_{1,N}} g_N(W), \sup_{W\in\Omega_N\cap\Omega_{1,N}^C} g_N(W)\right) \\ &\leq \sup_{W\in\Omega_N\cap\Omega_{1,N}} g_N(W) + \sup_{W\in\Omega_N\cap\Omega_{1,N}^C} g_N(W).\end{aligned}$$

Thus, setting $g_N(W) = \sum_{i=1}^{K_n}(\sum_{j=1}^{i} w_j)^2|u'(P_{i+1} - P_i)f/N|/S_\lambda(W)$ and using (A.19), (A.20), and (A.22), one obtains the bound

$$\sup_{W\in\Omega} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)} \tag{A.23}$$
$$\leq \frac{C(u'(P_{K_n+1}-P_1)u/N)^{1/2}}{\sqrt{\omega}N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}} + \sqrt{\omega}N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}\frac{C(u'(P_{K_n+1}-P_1)u/N)^{1/2}}{\inf_{W\in\Omega} S_\lambda(W)}.$$

It then follows that for any $\eta_1 > 0$ that

$$\Pr\left[\left|\sup_{W\in\Omega} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1}-P_i)f/N|}{S_\lambda(W)}\right| > \eta_1 \middle| z\right]$$
$$\leq \frac{1}{\sqrt{\omega}}\frac{C(E[u'(P_{K_n+1}-P_1)u/N|z])^{1/2}}{N^{(-\alpha+\varepsilon/2)/(2\alpha+1)}} + \frac{(E[u'(P_{K_n+1}-P_1)u/N|z])^{1/2}}{N^{-2\alpha/(2\alpha+1)}}O_p\big(N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}\big),$$

where the inequality uses Markov's inequality, (A.23), and Lemma A.8. Next, note that

$$\frac{C(E[u'(P_{K_n+1}-P_1)u/N|z])^{1/2}}{N^{(-\alpha+\varepsilon/2)/(2\alpha+1)}} \tag{A.24}$$
$$= \frac{1}{\sqrt{\omega}}\frac{C\sqrt{(K_{n+1}-1)/N}}{N^{(-\alpha+\varepsilon/2)/(2\alpha+1)}}$$
$$= o\big(N^{(-\varepsilon/2-\varepsilon/2)/(2\alpha+1)}\big) = o(1)$$

and

$$E\big[u'P_{K_{n+1}}u/N|z\big]^{1/2}O_p\big(N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}\big)$$
$$= O_p\big(K_n^{1/2}N^{(-\alpha+\varepsilon/4)/(2\alpha+1)-1/2}\big)$$
$$= O_p\big(N^{(-2\alpha-\varepsilon/4)/(2\alpha+1)}\big) = o_p\big(N^{-2\alpha/(2\alpha+1)}\big)$$

such that

$$\frac{(E[u'(P_{K_n+1} - P_1)u/N|z])^{1/2}}{N^{-2\alpha/(2\alpha+1)}} O_p\big(N^{(-\alpha+\varepsilon/4)/(2\alpha+1)}\big) = o_p(1). \tag{A.25}$$

Using (A.24) and (A.25) then establishes that

$$\Pr\left[\left|\sup_{W\in\Omega} \frac{\sum_{i=1}^{K_n}\left(\sum_{j=1}^{i} w_j\right)^2 |u'(P_{i+1} - P_i)f/N|}{S_\lambda(W)}\right| > \eta_1 \middle| z\right] = o(1) + o_p(1).$$

This completes the proof of the lemma. *Q.E.D.*

A.2. *Proofs of Theorems and Corollaries*

PROOF OF THEOREM A.1: The MA2SLS estimator has the form

$$\sqrt{N}(\hat\beta - \beta_0) = \hat H^{-1}\hat h, \quad \hat H = X'P(W)X/N,$$
$$\hat h = X'P(W)\epsilon/\sqrt{N}.$$

Also $\hat H$ and $\hat h$ are decomposed as

$$\hat h = h + T_1^h + T_2^h,$$
$$T_1^h = -f'(I - P(W))\epsilon/\sqrt{N}, \quad T_2^h = u'P(W)\epsilon/\sqrt{N},$$
$$\hat H = H + T_1^H + T_2^H + T_3^H + Z^H,$$
$$T_1^H = -f'(I - P(W))f/N, \quad T_2^H = (u'f + f'u)/N,$$
$$T_3^H = u'P(W)u/N,$$
$$Z^H = \big(u'(I - P(W))f + f'(I - P(W))u\big)/N.$$

We show that the conditions of Lemma A.1 are satisfied and $S(W)$ has the form given in the theorem. Let $\rho_{W,N} = \mathrm{tr}(S(W))$. Differently from Donald and Newey (2001), we extend the MA2SLS to order $K'W/N$. It is important to point out that since W can contain negative weights, it is possible that $(K'W)^2/N$ is not the dominating term in $S(W)$. For example, $K'W = 0$ is allowed. However, $K'W/N = O(S(W))$ by construction.

Now $h = O_p(1)$ and $H = O_p(1)$ by Lemma A.5(vi). As

$$T^h = T_1^h + T_2^h = -f'(I - P(W))\epsilon/\sqrt{N} + u'P(W)\epsilon/\sqrt{N},$$

Lemma A.6(ii) and (iii) implies that

$$T_1^h = O_p\big(\Delta(W)^{1/2}\big)$$

and

$$T_2^h = O_p\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/\sqrt{N}\right), \tag{A.26}$$

so

$$T^h = O_p\big(\Delta(W)^{1/2}\big) + O_p\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/\sqrt{N}\right),$$

where $\Delta(W) = o_p(1)$ by Lemma A.6(i), $K'W/\sqrt{N} = o(1)$ by $|K'W|/\sqrt{N} \le K'W^+/\sqrt{N} = o(1)$, $\sum_i (P_{ii}(W))^2 = o_p(K'W^+)$ by Lemma A.5(ii), and $W'\Gamma W = O(K'W^+)$ by Lemma A.6(xii). Therefore $T^h = o_p(1)$. Next, we observe $T_1^H = O(\Xi(W))$ by the definition. Lemmas A.6(i) and A.7 imply that $T_1^H = o_p(1)$; $T_2^H = O_p(1/\sqrt{N})$ by the central limit theorem (CLT). A similar argument for T_2^h implies

$$T_3^H = O_p\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/N\right). \tag{A.27}$$

Now we analyze

$$\|T_1^h\| \cdot \|T_1^H\| = O_p\big(\Delta(W)^{1/2}\Xi(W)\big) = o_p(\rho_{W,N})$$

by Lemma A.7. It holds that

$$\|T_1^h\| \cdot \|T_2^H\| = O_p\big(\Delta(W)^{1/2}/\sqrt{N}\big) = o_p(\rho_{W,N})$$

because, by Lemma A.6(vi), one can take $g(W) = N(\mathrm{tr}(S(W)) - \Delta(W))$. From Lemma A.3, it follows that $W'\Gamma W \to \infty$ as $N \to \infty$. This implies that $g(W) \to \infty$. Then, by Lemma A.6(vi), it follows that

$$\Delta(W)^{1/2}/\sqrt{N} = o_p\left(\frac{g(W)}{N} + \Delta(W)\right) = o_p\big(\mathrm{tr}(S(W))\big) = o_p(\rho_{W,N}).$$

Next,

$$\|T_1^h\| \cdot \|T_3^H\| = O_p\left(\Delta(W)^{1/2} \max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/N\right)$$

$$= o_p\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/ N\right)$$

$$= o_p(\rho_{W,N})$$

by Lemma A.6(i), (A.27), and the fact (as noted before) that $T_3^H = O(\operatorname{tr}(S(W)))$. Next, (A.26) and the definition of T_1^H imply that

$$\|T_2^h\| \cdot \|T_1^H\|$$

$$= O_p\left(\Xi(W) \max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/ \sqrt{N}\right)$$

$$= o_p\left(\Delta(W)^{1/2} \max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/ \sqrt{N}\right)$$

by Lemma A.7. By similar arguments as before, it follows from Lemma A.6(vi) that

$$\Delta(W)^{1/2}|K'W|/\sqrt{N} \leq (K'W)^2/N + \Delta(W) = O(\rho_{W,N})$$

and $\Delta(W)^{1/2} = o_p(1)$ such that $o_p(\Delta(W)^{1/2}K'W/\sqrt{N}) = o_p(\rho_{W,N})$ as required. Lemma A.6(vi) gives

$$\Delta(W)^{1/2} \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\Big/ \sqrt{N}$$

$$= O_p\left(\frac{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}{N} + \Delta(W)\right) = O_p(\rho_{W,N}).$$

Thus, we have $\|T_2^h\| \cdot \|T_1^H\| = o_p(\rho_{W,N})$. From (A.26) it follows that

$$\|T_2^h\| \cdot \|T_2^H\| = O_p\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/ N\right),$$

where $K'W/N = O(\operatorname{tr}(S(W)))$ and $\sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}/N = o_p(\operatorname{tr}(S(W)))$. By (A.26) and (A.27), it follows that

$$\|T_2^h\| \cdot \|T_3^H\|$$

$$= O_p\left(\max\left(|K'W|^2, \left((W'\Gamma W) + \sum_i (P_{ii}(W))^2\right)\right)\Big/ N^{3/2}\right)$$

$$= o_p(\rho_{W,N})$$

because $(|K'W|/N)^{3/2} = o(\rho_{W,N})$ and $((W'\Gamma W) + \sum_i (P_{ii}(W))^2)/N = O_p(\rho_{W,N})$. Similarly, $\|T_2^h\|^2\|T_1^H\| = o_p(\rho_{W,N})$, $\|T_2^h\|^2\|T_2^H\| = o_p(\rho_{W,N})$, and $\|T_2^h\|^2\|T_3^H\| = o_p(\rho_{W,N})$. For $\|T^H\|^2$, we have

$$\|T_1^H\|^2 = O_p(\Xi(W)^2) = o_p(\rho_{W,N}) \quad \text{by Lemma A.7,}$$

$$\|T_2^H\|^2 = O_p(1/N) = o_p(\rho_{W,N}),$$

$$\|T_3^H\|^2 = O_p\left(\left(\max\left(|K'W|, \sqrt{(W'\Gamma W) + \sum_i (P_{ii}(W))^2}\right)\Big/N\right)^2\right) = o_p(\rho_{W,N}),$$

so that by the Cauchy–Schwarz inequality, $\|T^H\|^2 = o_p(\rho_{W,N})$.

As $\|Z^h\| = 0$ in our case, $\|Z^h\| = o_p(\rho_{W,N})$. The last part, which we need to show being equal to $o_p(\rho_{W,N})$, is $\|Z^H\|$. Now $Z^H = u'(I - P(W))f/N + f'(I - P(W))u/N$ and both terms are $O_p(\Delta(W)^{1/2}/\sqrt{N}) = o_p(g(W)/N + \Delta(W)) = o_p(\rho_{W,N})$ for $g(W) = N(\mathrm{tr}(S(W)) - \Delta(W))$ by Lemma A.6(vi). Therefore, we have $\|Z^H\| = o_p(\rho_{W,N})$.

Note that we have shown $\hat{H} = H + o_p(1)$ and $\hat{h} = h + o_p(1)$. Lemma A.1 can now be applied, where the discussion above indicates

$$\begin{aligned} Z^A(W) = {} & -hT_1^{h\prime}H^{-1}\left(\sum_{j=1}^3 T_j^H\right)' - \left(\sum_{j=1}^3 T_j^H\right)H^{-1}T_1^h h' \\ & - T_1^h h' H^{-1}\left(\sum_{j=1}^3 T_j^H\right)' - \left(\sum_{j=1}^3 T_j^H\right)H^{-1}hT_1^{h\prime} \\ & - hT_2^{h\prime}H^{-1}T_3^{H\prime} - T_3^H H^{-1}T_2^h h' \\ & - T_2^h h' H^{-1}T_3^{H\prime} - T_3^H H^{-1}hT_2^{h\prime} \\ & - (T_1^h + T_2^h)(T_1^h + T_2^h)'H^{-1}\left(\sum_{j=1}^3 T_j^H\right)' \\ & - \left(\sum_{j=1}^3 T_j^H\right)H^{-1}(T_1^h + T_2^h)(T_1^h + T_2^h)' \\ = {} & o_p(\rho_{W,N}) \end{aligned}$$

and

$$
\begin{aligned}
\hat{A}(W) = {} & (h + T_1^h + T_2^h)(h + T_1^h + T_2^h)' \\
& - hh'H^{-1}\left(\sum_{j=1}^{3} T_j^H\right)' - \left(\sum_{j=1}^{3} T_j^H\right)H^{-1}hh' \\
& - hT_2^{h\prime}H^{-1}(T_1^H + T_2^H)' - (T_1^H + T_2^H)H^{-1}T_2^h h' \\
& - T_2^h h'H^{-1}(T_1^H + T_2^H)' - (T_1^H + T_2^H)H^{-1}hT_2^{h\prime}.
\end{aligned}
$$

Now we calculate the expectation of each term in $\hat{A}(W)$. First of all, $E[hh'|z] = E[f'\epsilon\epsilon' f|z] = \sigma_\epsilon^2 H$. Second, $E[hT_1^{h\prime}|z] = E[-f'\epsilon\epsilon'(I - P(W))f/N|z] = -\sigma_\epsilon^2 f'(I - P(W))f/N$. Similarly, $E[T_1^h h'|z] = -\sigma_\epsilon^2 f'(I - P(W))f/N$. Third,

$$
\begin{aligned}
E[hT_2^{h\prime}|z] &= E[f'\epsilon\epsilon' P(W)u/N|z] \\
&= E[\epsilon_1^2 u_1]\sum_i f_i' P_{ii}(W)/N = O_p(K'W^+/N)
\end{aligned}
$$

by Lemma A.6(v). This implies that $E[T_2^h h'|Z] = O_p(K'W/N)$ too. Fourth,

$$
\begin{aligned}
& E[T_1^h T_1^{h\prime}|z] \\
& \quad = E\left[\frac{f'(I - P(W))\epsilon\epsilon'(I - P(W))f}{N}\Big| z\right] \\
& \quad = \sigma_\epsilon^2 \frac{f'(I - P(W))(I - P(W))f}{N}.
\end{aligned}
$$

Fifth,

$$
\begin{aligned}
& E[T_1^h T_2^{h\prime}|z] \\
& \quad = -E\big[f'(I - P(W))\epsilon\epsilon' P(W)u/N|z\big] \\
& \quad = -f'(I - P(W))\mu(W)/N
\end{aligned}
$$

by Lemma A.6(viii). Again, we have $E[T_2^h T_1^{h\prime}|z] = -\mu(W)'(I - P(W))f/N$. Sixth,

$$
\begin{aligned}
E[T_2^h T_2^{h\prime}|z] &= E\left[\frac{u'P(W)\epsilon\epsilon' P(W)u}{N}\Big| z\right] \\
&= \sigma_{u\epsilon}\sigma_{u\epsilon}' \frac{(K'W)^2}{N} + (\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')\frac{(W'\Gamma W)}{N} \\
&\quad + \mathrm{Cum}[\epsilon_i, \epsilon_i, u_i, u_i'] \sum_i (P_{ii}(W))^2
\end{aligned}
$$

by Lemma A.6(iv). Seventh,

$$E[hh'H^{-1}T_1^H|z] = -E\left[\frac{f'\epsilon\epsilon' f H^{-1} f'(I-P(W))f}{N^2}\Big| z\right]$$
$$= -\sigma_\epsilon^2 \frac{f'(I-P(W))f}{N}.$$

Also, we have $E[T_1^H H^{-1}hh'|Z] = -\sigma_\epsilon^2 f'(I-P(W))f/N$. Lemma A.6(vii) implies

$$E[hh'H^{-1}T_2^H|z] = E\left[\frac{hh'H^{-1}(u'f+f'u)}{N}\Big| z\right]$$
$$= O_p\left(\frac{1}{N}\right)$$

and $E[T_2^H H^{-1}hh'|z] = O_P(1/N)$. Also,

$$E[hh'H^{-1}T_3^H|z] = E\left[\frac{f'\epsilon\epsilon' f H^{-1}u'P(W)u}{N^2}\Big| z\right]$$
$$= \sigma_\epsilon^2 \Sigma_u \frac{K'W}{N} + O_p\left(\frac{1}{N}\right)$$

by Lemma A.6(ix). Next,

$$\begin{aligned}
&E[hT_2^{h\prime}H^{-1}T_1^H|z]\\
&\quad = -E\left[\frac{f'\epsilon\epsilon' P(W)uH^{-1}f'(I-P(W))f}{N^2}\Big| z\right]\\
&\quad = \frac{1}{N}\sum_i f_i P_{ii}(W)E[\epsilon_i^2 u_i']H^{-1}\frac{f'(I-P(W))f}{N}\\
&\quad = O_p((K'W^+/N)\Xi(W))\\
&\quad = o_p(\rho_{W,N})
\end{aligned}$$

by Lemma A.6(v) and

$$\begin{aligned}
&E[hT_2^{h\prime}H^{-1}T_2^H|z]\\
&\quad = E\left[\frac{f'\epsilon\epsilon' P(W)uH^{-1}(u'f+f'u)}{N^2}\Big| z\right]\\
&\quad = O_p\left(\frac{1}{N}\right) + \frac{K'W}{N}\left(\frac{1}{N}\sum_i f_i\sigma_{u\epsilon}' H^{-1}\sigma_{u\epsilon} f_i' + \frac{1}{N}\sum_i f_i\sigma_{u\epsilon}' H^{-1} f_i \sigma_{u\epsilon}'\right)
\end{aligned}$$

by Lemma A.6(x). Similarly, it follows that

$$\begin{aligned}
&E[T_2^h h' H^{-1} T_2^H | z] \\
&= E\left[\frac{u' P(W) \epsilon \epsilon' f H^{-1}(u' f + f' u)}{N^2} \Big| z\right] \\
&= O_p\left(\frac{1}{N}\right) + \frac{K'W}{N}\left(d \sigma_{u\epsilon} \sigma_{u\epsilon}' + \sigma_{u\epsilon} \frac{1}{N} \sum_i f_i' H^{-1} \sigma_{u\epsilon} f_i'\right).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&E[\hat{A}(K)|z] \\
&= \sigma_\epsilon^2 H - 2\sigma_\epsilon^2 \frac{f'(I - P(W))f}{N} + \sigma_\epsilon^2 \frac{f'(I - P(W))(I - P(W))f}{N} \\
&\quad + E[\epsilon_1^2 u_1] \sum_i f_i' P_{ii}(W)/N + \sum_i f_i P_{ii}(W) E[\epsilon_1^2 u_1']/N \\
&\quad + \frac{f'(I - P(W))\mu(W)}{N} + \frac{\mu(W)'(I - P(W))f}{N} + \sigma_{u\epsilon} \sigma_{u\epsilon}' \frac{(K'W)^2}{N} \\
&\quad + (\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon} \sigma_{u\epsilon}') \frac{(W' \Gamma W)}{N} + o_p\left(\frac{K'W}{N}\right) \\
&\quad + 2\sigma_\epsilon^2 \frac{f'(I - P(W))f}{N} + O_p\left(\frac{1}{N}\right) - 2\sigma_\epsilon^2 \Sigma_u \frac{K'W}{N} \\
&\quad - \frac{K'W}{N} 2\left(d \sigma_{u\epsilon} \sigma_{u\epsilon}' + \frac{1}{N} \sum_i f_i \sigma_{u\epsilon}' H^{-1} \sigma_{u\epsilon} f_i' \right. \\
&\quad \left. + \frac{1}{N} \sum_i (f_i \sigma_{u\epsilon}' H^{-1} f_i \sigma_{u\epsilon}' + \sigma_{u\epsilon} f_i' H^{-1} \sigma_{u\epsilon} f_i')\right) \\
&\quad + o_p(\rho_{W,N}) \\
&= \sigma_\epsilon^2 H + \sigma_\epsilon^2 \frac{f'(I - P(W))(I - P(W))f}{N} \\
&\quad + E[\epsilon_1^2 u_1] \sum_i f_i' P_{ii}(W)/N + \sum_i f_i P_{ii}(W) E[\epsilon_1^2 u_1']/N \\
&\quad + \frac{f'(I - P(W))\mu(W)}{N} + \frac{\mu(W)'(I - P(W))f}{N} \\
&\quad + \sigma_{u\epsilon} \sigma_{u\epsilon}' \frac{(K'W)^2}{N} + (\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon} \sigma_{u\epsilon}') \frac{(W' \Gamma W)}{N}
\end{aligned}$$

$$-2\frac{K'W}{N}\left(\sigma_\epsilon^2\Sigma_u + d\sigma_{u\epsilon}\sigma_{u\epsilon}' + \frac{1}{N}\sum_i f_i\sigma_{u\epsilon}'H^{-1}\sigma_{u\epsilon}f_i' \right.$$
$$\left. + \frac{1}{N}\sum_i (f_i\sigma_{u\epsilon}'H^{-1}f_i\sigma_{u\epsilon}' + \sigma_{u\epsilon}f_i'H^{-1}\sigma_{u\epsilon}f_i')\right)$$
$$+ o_p(\rho_{W,N}),$$

where the last equality holds because $1/N = o_p(\rho_{W,N})$ and $o_p((\Delta(W) \times K'W/N)^{1/2}) = o_p(\rho_{W,N})$ by the fact that $(\Delta(W)K'W/N)^{1/2} \leq K'W/N + \Delta(W)$. *Q.E.D.*

We omit the proofs of Corollaries A.1 and A.3 because they are trivial given Theorem A.1.

PROOF OF COROLLARY A.2: We note that in this case $K'W = K'W^+$. Thus, $\sum_i (P_{ii}(W))^2 = o_p(K'W)$ by Lemma A.5(ii) and $f'Q(W)\mu(W)/N = o_p(K'W/N + \Delta(W))$ by Lemma A.6 (viii). Therefore, we have equation (A.3).

To derive equation (A.4), we note that

$$W'\Gamma W = \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j \min(i,j) \leq \sum_{i=1}^{M}\sum_{j=1}^{M} w_i w_j j = \sum_{i=1}^{M} w_i \sum_{j=1}^{M} w_j j$$
$$= W'\mathbf{1}_M K'W = K'W,$$

which means $W'\Gamma W = O(K'W)$. Moreover, $\sum_{i=1}^{N} f_i P_{ii}(W) = O_p(K'W)$ by Lemma A.6(v). Therefore, we have equation (A.4). *Q.E.D.*

PROOF OF THEOREM A.3: The result is established by constructing a sequence in Ω_P that dominates the optimal choice in Ω_{sq}. By Corollary A.2, the formula of $S_\lambda(W)$ for MA2SLS when $W \in \Omega_P$ is

$$A\frac{(K'W)^2}{N} + \sigma_\epsilon^2 \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} w_j w_i \gamma_{\max(i,j)}$$

with $A = \|\lambda' H^{-1}\sigma_{u\epsilon}\|^2$ (the other two terms in (A.2) can be ignored). Let M_{sq} be the optimal number of instruments picked by the Donald and Newey (2001) algorithm. For $a \in (0,1)$, let $M_1 = (1-a)M_{sq}$ and $M_2 = (1+a)M_{sq}$, and choose W^* such that it has only two nonzero elements $w_{M_1} = w_{M_2} = 0.5$. Then $K'W^* = M_{sq}$ and

$$\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} w_j w_i \gamma_{\max(i,j)} = 0.25\gamma_{M_1} + 0.75\gamma_{M_2}.$$

Then

$$\begin{aligned}
&\frac{\min_{W\in\Omega_P} S_\lambda(W)}{\min_{W\in\Omega_{sq}} S_\lambda(W)} \\
&\quad\le \frac{S_\lambda(W^*)}{S_{sq}(M_{sq})} \\
&\quad= \frac{A(K'W^*)^2/N + \sigma_\epsilon^2 0.25\gamma_{M_1} + \sigma_\epsilon^2 0.75\gamma_{M_2}}{A(M_{sq})^2/N + \sigma_\epsilon^2\gamma_{M_{sq}}} \\
&\quad= \frac{A(M_{sq})^2/(N\sigma_\epsilon^2\gamma_{M_{sq}}) + 0.25(\gamma_{M_1}/\gamma_{M_{sq}}) + 0.75(\gamma_{M_2}/\gamma_{M_{sq}})}{A(M_{sq})^2/(N\sigma_\epsilon^2\gamma_{M_{sq}}) + 1},
\end{aligned}$$

where $\gamma = \limsup_{N\to\infty} A(M_{sq})^2/(N\gamma_{M_{sq}}) < \infty$ because M_{sq} sets the rates of the bias and the variance to be equal. The above expression is bounded by 1 if

$$0.25\big(\gamma_{M_1}/\gamma_{M_{sq}}\big) + 0.75\big(\gamma_{M_2}/\gamma_{M_{sq}}\big) < 1.$$

By assumption, for N large enough, it follows that, with probability close to 1,

$$\begin{aligned}
&0.25\big(\gamma_{M_1}/\gamma_{M_{sq}}\big) + 0.75\big(\gamma_{M_2}/\gamma_{M_{sq}}\big) \\
&\quad= 0.25(1-a)^{-2\alpha} + 0.75(1+a)^{-2\alpha} + o(|a|^{2\alpha}).
\end{aligned}$$

Consider the function

$$h(a) = 0.25(1-a)^{-2\alpha} + 0.75(1+a)^{-2\alpha},$$

where $h(0) = 1$, $\partial h(a)/\partial a = 0.5\alpha(1-a)^{-2\alpha-1} - 1.5\alpha(1+a)^{-2\alpha-1}$ such that $\partial h(0)/\partial a = -1\alpha$. This implies that for some a, possibly close to zero, $h(a) < 1$ and thus $0.25(\gamma_{M_1}/\gamma_{M_{sq}}) + 0.75(\gamma_{M_2}/\gamma_{M_{sq}}) < 1$.

When $W \in \Omega_B$, the formula of $S_\lambda(W)$ for MA2SLS is

$$S_\lambda(W) = A\frac{(W'\Gamma W)}{N} + \sigma_\epsilon^2 \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} w_j w_i \gamma_{\max(i,j)},$$

where $A = \lambda' H^{-1}(\sigma_\epsilon^2 \Sigma_u + \sigma_{u\epsilon}\sigma_{u\epsilon}')H^{-1}\lambda$ while the MSE for the Nagar estimator with M instruments is $AM/(N-M) + \sigma_\epsilon^2\gamma_M$. Let M_N be the choice of M that minimizes $S_\lambda(W)$ when $W = W_N$ as defined in Remark A.1. For $a \in (0,1)$, let $M_1 = (1-a)M_N$ and $M_2 = (1+a)M_N$. Define $w^* = N/(N-M_N)$ and choose W^* such that W^* has only three nonzero elements $w_{M_1} = w_{M_2} = 1/2w^*$ and $w_N = -M_N/(N-M_N)$. For brevity write w_1 and w_2 instead of w_{M_1} and w_{M_2}. Then $w_1 + w_2 + w_N = 1$ and $K'W^* = 0$ such that $W^* \in \Omega_B$. Note that $W_N'\Gamma W_N = ((w^*)^2 + 2w^* w_N)M_N + w_N^2 N = M_N N/(N - M_N)$ and

$$\begin{aligned} W^{*\prime}\Gamma W^* &= w_1^2 M_1 + w_2^2 M_2 + 2w_1 w_2 M_1 \\ &\quad + w_N^2 N + 2w_N(w_1 M_1 + w_2 M_2) \\ &= w_1^2 M_1 + w_2^2 M_2 + 2w_1 w_2 M_1 + w_N^2 N + 2w_N w^* M_N \\ &= ((w^*)^2 + 2w_N w^*)M_N + w_N^2 N - (1/2)(w^*)^2 a M_N \end{aligned}$$

such that $W^{*\prime}\Gamma W^* < W_N'\Gamma W_N$. In the same way it follows that, for W^*,

$$\begin{aligned} \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} w_j w_i \gamma_{\max(i,j)} &= w_1^2 \gamma_{M_1} + (w_2^2 + 2w_1 w_2)\gamma_{M_2} \\ &\quad + (w_N^2 + 2w_N(w_1 + w_2))\gamma_N \\ &= (w^*)^2(\gamma_{M_1}/4 + 3\gamma_{M_2}/4) + (w_N^2 + 2w_N w^*)\gamma_N. \end{aligned}$$

Since the term $(w_N^2 + 2w_N w^*)\gamma_N$ is of smaller order than $S_\lambda(W_N)$, the result now follows if $(\gamma_{M_1}/4 + 3\gamma_{M_2}/4)/\gamma_{M_N} \le 1$ wpa1. But this follows from the same arguments as for the proof of the first statement of the theorem.

For MALIML, the formula of $S_\lambda(W)$ is

$$S_\lambda(W) = A\frac{(W'\Gamma W)}{N} + \sigma_\epsilon^2 \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} w_j w_i \gamma_{\max(i,j)},$$

where $A = \lambda' H^{-1}(\sigma_\epsilon^2 \Sigma_u - \sigma_{u\epsilon}\sigma_{u\epsilon}')H^{-1}\lambda$. Let M_{sq} be the optimal number of instruments chosen by the Donald–Newey method. The MSE of the estimator that uses M_{sq} instruments is $AM_{sq}/N + \sigma_\epsilon^2 \gamma_{M_{sq}}$. For $a \in (0,1)$, let $M_1 = (1 - a)M_{sq}$ and $M_2 = (1+a)M_{sq}$, and choose W^* such that it has only two nonzero elements $w_{M_1} = w_{M_2} = 0.5$. The MSE of the estimator with W^* is

$$\frac{A}{N}(0.75M_1 + 0.25M_2) + \sigma_\epsilon^2(0.25\gamma_{M_1} + 0.75\gamma_{M_2}).$$

We note that $0.75M_1 + 0.25M_2 = M_{sq} - 0.5aM_{sq} < M_{sq}$. Moreover, we have $0.25\gamma_{M_1} + 0.75\gamma_{M_2} < \gamma_{M_{sq}}$ by the same arguments as for the proof of the first statement of the theorem. Therefore, the desired result is shown. *Q.E.D.*

PROOF OF THEOREM A.4: We follow the proof of Donald and Newey (2001, Lemma A9). We first consider the case for $S(W)$ defined in (A.2) and $\hat{S}_\lambda(W)$ defined in (2.5). Note that when $\Omega = \Omega_U$ and $\Omega = \Omega_B$, the optimal weight, W^*, is well defined and has a closed form (see the discussion in Section A.5). When $\Omega = \Omega_C$ or Ω_P, we note that $S_\lambda(W)$ is continuous in W and Ω is a compact set, which implies that the optimal weight, W^*, is well defined in this case too. Thus $\inf_{W\in\Omega} S_\lambda(W) = S_\lambda(W^*)$ for some $W^* \in \Omega$ holds. It then follows that

$$0 \le 1 - \frac{\inf\limits_{W\in\Omega} S_\lambda(W)}{S_\lambda(\hat{W})} \le 4 \sup_{W\in\Omega}\left|\frac{\hat{S}_\lambda(W)}{S_\lambda(W)} - 1\right|.$$

The result now follows from Lemma A.10.

Next, we consider the case for $S(W)$ defined in (A.6) and $\hat{S}_\lambda(W)$ defined in (2.7) (the case for MALIML). We follow the steps taken in the above argument. First, we show that $\inf_{W\in\Omega} S_\lambda(W) = O_p(N^{-2\alpha/(2\alpha+1)})$. The weighting vector, $\tilde{W}$, where $w_M = 1$ and $w_j = 0$ for $j \neq M$ for $M = O(N^{1/(2\alpha+1)})$, gives $S_\lambda(\tilde{W}) = O_p(N^{-2\alpha/(2\alpha+1)})$. The proof that this rate is sharp is exactly equivalent to the corresponding part of the proof of Lemma A.8. We then show that $\sup_{W\in\Omega}(\tilde{S}_\lambda(W)/S_\lambda(W)) - 1 = o_p(1)$, where

$$\begin{aligned}\tilde{S}_\lambda(W) = \lambda'\hat{H}^{-1}\Bigg((\hat{\sigma}_\epsilon^2\hat{\sigma}_\lambda^2 - \hat{\sigma}_{\lambda\epsilon}^2)\frac{W'\Gamma W}{N} \\ + \hat{\sigma}_\epsilon^2\frac{f'(I-P(W))(I-P(W))f}{N}\Bigg)\hat{H}^{-1}\lambda.\end{aligned}$$

This can be shown by following the same argument as that for the Ω_2 part of the proof of Lemma A.9. Last, we show that $\sup_{W\in\Omega}(\hat{S}_\lambda(W)/S_\lambda(W)) - 1 = o_p(1)$. The proof of this statement is the same as that of Lemma A.10. We then obtain the desired result. *Q.E.D.*

PROOF OF THEOREM A.5: Since it is easy to see that $X'X/N \to_p (E(f_i^2) + \sigma_u^2)$, we need to show

$$\frac{1}{N}X'P(W)X \to_p E(f_i^2) \tag{A.28}$$

and

$$\frac{1}{N}X'P(W)P(W)X \to_p E(f_i^2) \tag{A.29}$$

to obtain the desired result.

We have the decomposition

$$\begin{aligned}\frac{1}{N}X'P(W)X = \frac{1}{N}f'f - \frac{1}{N}f'(I-P(W))f \\ + \frac{1}{N}f'P(W)u + \frac{1}{N}u'P(W)f + \frac{1}{N}u'P(W)u.\end{aligned}$$

By Lemmas A.6(i) and A.7, it holds that

$$\frac{1}{N}f'(I-P(W))f = o_p(1).$$

Since

$$\frac{1}{N}f'P(W)u = \frac{1}{N}f'u - \frac{1}{N}f'(I-P(W))u,$$

Lemmas A.5(vi) and A.6(ii) (by replacing ϵ by u) imply that

$$\frac{1}{N}f'P(W)u = o_p(1).$$

Similarly, it follows that $u'P(W)f/N = o_p(1)$. Last, Lemma A.6(iii) and Assumption 4 imply that

$$\frac{1}{N}u'P(W)u = o_p(1).$$

Thus, we have shown (A.28).

We now consider (A.29). We have the decomposition

$$\begin{aligned}
&\frac{1}{N}X'P(W)P(W)X \\
&\quad = \frac{1}{N}f'f - \frac{1}{N}f'(I-P(W)P(W))f \\
&\qquad + \frac{1}{N}f'P(W)P(W)u + \frac{1}{N}u'P(W)P(W)f + \frac{1}{N}u'P(W)P(W)u.
\end{aligned}$$

We have that

$$\begin{aligned}
\frac{1}{N}f'(I-P(W)P(W))f &= \sum_{s_1=1}^{M}\sum_{s_2=1}^{M} w_{s_1}w_{s_2}f'\big(I-P_{\min(s_1,s_2)}\big)f \\
&= \sum_{j=1}^{M}\left(2w_j\left(\sum_{s=j+1}^{M} w_s\right)+w_j^2\right)\tilde{\gamma}_j,
\end{aligned}$$

where $\tilde{\gamma}_j = f'(I-P_j)f/N$. It follows that

$$\sum_{j=1}^{M}\left(2w_j\left(\sum_{s=j+1}^{M} w_s\right)+w_j^2\right)\tilde{\gamma}_j = \sum_{j=1}^{M} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\tilde{\gamma}_j.$$

Take L such that $L \to \infty$. We have that

$$\begin{aligned}
&\left|\sum_{j=1}^{M} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\tilde{\gamma}_j\right| \\
&\quad \le \left|\sum_{j=1}^{L} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\tilde{\gamma}_j\right| + \left|\sum_{j=L+1}^{M} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\right|\tilde{\gamma}_L \\
&\quad = \left|\sum_{j=1}^{L} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\tilde{\gamma}_j\right| + o_p(1)
\end{aligned}$$

since $\tilde{\gamma}_L = o_p(1)$ and $W \in l_1$ implies that $|\sum_{j=L+1}^{M} w_j(2-2\sum_{s=1}^{j} w_s + w_j)|$ is bounded. Then, since $\sum_{j=1}^{L} |w_j| = o(1)$ by assumption, we have

$$\left|\sum_{j=1}^{L} w_j\left(2-2\sum_{s=1}^{j} w_s + w_j\right)\tilde{\gamma}_j\right| = o_p(1).$$

It follows that

$$\frac{1}{N} f'(I - P(W)P(W))f = o_p(1).$$

We have that

$$E\left(\frac{1}{N} f'P(W)P(W)u\right) = 0$$

and

$$\begin{aligned}
&E\left(\left(\frac{1}{N} f'P(W)P(W)u\right)^2\right) \\
&\quad = E\left(\frac{1}{N^2} f'P(W)P(W)uu'P(W)P(W)f\right) \\
&\quad = \frac{1}{N^2}\sigma_u^2 f'P(W)P(W)P(W)P(W)f \\
&\quad = \frac{1}{N^2}\sigma_u^2 \sum_{s_1,s_2,s_3,s_4=1}^{M} w_{s_1}w_{s_2}w_{s_3}w_{s_4} f'P_{\min(s_1,s_2,s_3,s_4)}f \\
&\quad \le \frac{1}{N^2}\sigma_u^2 \sum_{s_1,s_2,s_3,s_4=1}^{M} |w_{s_1}w_{s_2}w_{s_3}w_{s_4}| f'f = o_p(1)
\end{aligned}$$

because $f'f/N = O_p(1)$ by Lemma A.5(vi) and $W \in l_1$ by Assumption 4. It therefore follows that

$$\frac{1}{N} f'P(W)P(W)u = o_p(1).$$

Similarly, we have that $u'P(W)P(W)f/N = o_p(1)$. Last, we observe that

$$E\left(\frac{1}{N} u'P(W)P(W)u\right) = \sigma_u^2 \frac{W'\Gamma W}{N}$$

by Lemma 1.2 of Hansen (2007). Assumption 4 and the Markov inequality imply that

$$\frac{1}{N} u'P(W)P(W)u = o_p(1).$$

Therefore, (A.29) is shown and we have obtained the desired result. *Q.E.D.*

A.3. *Lemmas for MALIML*

As the first step, we show the consistency of MALIML and derive its asymptotic distribution. Define the LIML estimator based on the first m instruments as

$$\hat{\beta}_{L,m} = \arg\min_{\beta} (y - X\beta)'P_m(y - X\beta)/((y - X\beta)'(y - X\beta)).$$

We first establish uniform consistency $\sup_{m \le M} |\hat{\beta}_{L,m} - \beta_0| \to_p 0$ for $M/N \to 0$. This result is then used to establish the uniform convergence of $\hat{\Lambda}(W)$ over M and W satisfying Assumption 5(ii).

LEMMA A.11: *If Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*, *then the following equalities hold*:

(i) $\sup_{m \le M} \epsilon' P_m \epsilon / N = o_p(1)$.
(ii) $\sup_{m \le M} f'(I - P_m)\epsilon / N = O_p(1/\sqrt{N})$.
(iii) $\sup_{m \le M} u' P_m \epsilon / N = o_p(1)$.

PROOF: For (i), we observe that

$$\sup_{k \le M} \epsilon' P_k \epsilon / N \le \epsilon' P_M \epsilon / N$$

and

$$E[\epsilon' P_M \epsilon | z] = \sigma_\epsilon^2 \operatorname{tr}(P_M) = \sigma_\epsilon^2 M$$

such that

$$\Pr\Big(\sup_{m\le M}|\epsilon' P_m\epsilon/N|>\eta\Big|z\Big)\le \Pr(|\epsilon' P_M\epsilon/N|>\eta|z)$$
$$\le \frac{1}{\eta N}E[\epsilon' P_M\epsilon|z]\to 0.$$

For part (ii), note that $E[f'(I-P_m)\epsilon|z]=0$ such that

$$\begin{aligned}&\sum_{m=1}^{M}\operatorname{tr}E[f'(I-P_m)\epsilon\epsilon(I-P_m)f/N|z]\\
&\quad\le \sup_{m\le M}\big(m^{2\alpha}\operatorname{tr}(f'(I-P_m)f)/N\big)\sigma_\epsilon^2\sum_{m=1}^{M}m^{-2\alpha}\\
&\quad=O_p(1),\end{aligned}$$

which shows that $\sup_{m\le M}f'(I-P_m)\epsilon/N=O_p(1/\sqrt{N})$.

For part (iii), note that $E[u'P_m\epsilon/N|z]=E[v'P_m\epsilon/N|z]+\sigma_{u\epsilon}/\sigma_\epsilon^2 E[\epsilon' P_m\epsilon/N|z]=0+\sigma_{u\epsilon}m/N$ and

$$\begin{aligned}
&E[\|u'P_m\epsilon/N-\sigma_{u\epsilon}m/N\|^2|z] \qquad \text{(A.30)}\\
&\quad\le M\max_{m\le M}\frac{E[\operatorname{tr}(u'P_m\epsilon\epsilon' P_m u-\sigma_{u\epsilon}\sigma_{u\epsilon}'m^2)|z]}{N^2}\\
&\quad=M\max_{m\le M}\left(\frac{\operatorname{tr}\Sigma_u\operatorname{tr}P_m}{N^2}+\sum_{i_1,\dots,i_4=1}^{N}\frac{\operatorname{tr}(E[u_{i_1}\epsilon_{i_3}]E[u_{i_4}'\epsilon_{i_2}])P_{m,i_1i_2}P_{m,i_3i_4}}{N^2}\right)\\
&\qquad+M\max_{m\le M}\left(\sum_{i=1}^{N}\frac{\operatorname{tr}(\operatorname{Cum}(u_i,u_i,\epsilon_i,\epsilon_i))P_{m,ii}^2}{N^2}\right)\\
&\quad\le\big(\operatorname{tr}\Sigma_u+\operatorname{tr}(\operatorname{Cum}(u_i,u_i,\epsilon_i,\epsilon_i))\big)\left(\frac{M}{N}\right)^2\\
&\qquad+M\max_{m\le M}\frac{\sigma_{u\epsilon}'\sigma_{u\epsilon}}{N^2}\sum_{i_1,i_2=1}^{N}P_{m,i_1i_2}P_{m,i_2i_1}\\
&\quad=o(1)+M\max_{m\le M}\left(\frac{\sigma_{u\epsilon}'\sigma_{u\epsilon}m}{N^2}\right)\\
&\quad=o(1),
\end{aligned}$$

where we used $P_{m,i_2 i_1} = P_{m,i_1 i_2}$ and $\sum_{i_1,i_2=1}^{N} P_{m,i_1 i_2} P_{m,i_2 i_1} = \sum_{i=1}^{N} P_{m,ii} = m$. Then

$$\begin{aligned}\|u'P_m\epsilon/N\| &\le \|u'P_m\epsilon/N - \sigma_{u\epsilon}m/N\| + \|\sigma_{u\epsilon}\|m/N \\ &\le \|u'P_m\epsilon/N - \sigma_{u\epsilon}m/N\| + \|\sigma_{u\epsilon}\|M/N \\ &= \|u'P_m\epsilon/N - \sigma_{u\epsilon}m/N\| + o(1),\end{aligned}$$

where the $o(1)$ term is uniform in $m \le M$. The result now follows from (A.30). *Q.E.D.*

LEMMA A.12: *If Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*, *then* $\sup_{m\le M} |\hat{\beta}_{L,m} - \beta_0| \to_p 0$.

PROOF: Define $\bar{X} \equiv (y, X)$ and $D_0 \equiv (\beta_0, I)$. $\bar{X}$ can be written as $\bar{X} = XD_0 + \epsilon e_1'$, where e_1 is the first unit (column) vector. Let $\hat{A}_m = \bar{X}'P_m\bar{X}/N$ and $A_m = D_0'\bar{H}_m D_0$. Let $\hat{B} = \bar{X}'\bar{X}/N$ and $B = E[\bar{X}_i\bar{X}_i']$ with $\bar{X}_i = (y_i, X_i)$.

Let $\tau = (1, -\beta')'$ and define the augmented parameter space $\overline{\Theta} = \{1\} \times \Theta$ such that $\tau \in \overline{\Theta}$. Then $(1, -\hat{\beta}_{L,m}')' = \arg\min_\tau \tau'\hat{A}_m\tau/(\tau'\hat{B}\tau)$. Essentially the same argument as that in the beginning of the proof of Lemma A.5 in Donald and Newey (2001) shows that $(1, -\beta_0')' = \arg\min_\tau \tau'A_m\tau/(\tau'B\tau)$. Then, letting $L_{n,m}(\tau) = \tau'\hat{A}_m\tau/(\tau'\hat{B}\tau)$ and $L_m(\tau) = \tau'A_m\tau/(\tau'B\tau)$, and noting that

$$\text{(A.31)}\quad \sup_{\tau\in\overline{\Theta}, m\le M} |L_{n,m}(\tau) - L_m(\tau)| \le \sup_{\tau\in\overline{\Theta}, m\le M}\left|\frac{\tau'(\hat{A}_m - A_m)\tau}{\tau'B\tau}\right| \sup_{\tau\in\overline{\Theta}}\left|\frac{\tau'B\tau}{\tau'\hat{B}\tau}\right| + \sup_{\tau\in\overline{\Theta}, m\le M}\left|\frac{\tau'A_m\tau}{\tau'B\tau}\right| \sup_{\tau\in\overline{\Theta}}\left|\frac{\tau'B\tau}{\tau'\hat{B}\tau} - 1\right|,$$

we note that $\tau'\hat{A}_m\tau/(\tau'\hat{B}\tau) \le 1$ uniformly in n, m, and τ. It follows that

$$\sup_{\tau\in\overline{\Theta}, m\le M}\left|\frac{\tau'A_m\tau}{\tau'B\tau}\right| \le 1 \quad \text{a.s.}$$

By a law of large numbers (LLN), $\hat{B} - B = o_p(1)$, which implies that $\sup_{\tau\in\overline{\Theta}} |\frac{\tau'B\tau}{\tau'\hat{B}\tau} - 1| = o_p(1)$. From Donald and Newey (2001, p. 1185) it follows that B is positive definite such that $\inf_\tau \tau'B\tau > \varepsilon > 0$ for some ε and

$$\text{(A.32)}\quad \sup_{\tau\in\overline{\Theta}, m\le M}\left|\frac{\tau'(\hat{A}_m - A_m)\tau}{\tau'B\tau}\right| \le \frac{\sup_{\tau\in\overline{\Theta}, m\le M} |\tau'(\hat{A}_m - A_m)\tau|}{\varepsilon}.$$

We now show that $\sup_{\tau\in\overline{\Theta},m\leq M}|\tau'(\hat{A}_m - A_m)\tau| = o_p(1)$. For this purpose, we observe that

$$\sup_{m\leq M}\frac{\sigma_\epsilon^2 m}{N} \leq \frac{\sigma_\epsilon^2 M}{N} = o(1), \tag{A.33}$$

$$\begin{aligned}
&\sup_{\tau\in\overline{\Theta},m\leq M}\frac{\tau' D_0' E[u'P_m u|z]D_0\tau}{N} \\
&\quad = \sup_{\tau\in\overline{\Theta},m\leq M}\frac{\mathrm{tr}(P_m E[uD_0\tau\tau' D_0' u'])}{N} \\
&\quad = \sup_{\tau\in\overline{\Theta},m\leq M}\frac{\mathrm{tr}(P_m)\tau' D_0'\Sigma_u D_0\tau}{N} \\
&\quad \leq \frac{M}{N}\sup_{\tau\in\overline{\Theta}}\tau' D_0'\Sigma_u D_0\tau = o(1),
\end{aligned} \tag{A.34}$$

where $\sup_{\tau\in\overline{\Theta}}\tau' D_0'\Sigma_u D_0\tau$ is bounded by Assumption 7, and

$$\begin{aligned}
&\sup_{\tau\in\overline{\Theta},m\leq M}\left|\frac{\tau' D_0' E[u'P_m\epsilon|z]e_1'\tau}{N}\right| \\
&\quad = \sup_{\tau\in\overline{\Theta},m\leq M}\left|\frac{\mathrm{tr}(P_m E[\epsilon\tau' D_0' u'])}{N}\right| \\
&\quad = \sup_{\tau\in\overline{\Theta},m\leq M}\left|\frac{\mathrm{tr}(P_m)\tau' D_0'\sigma_{u\epsilon}}{N}\right| \\
&\quad \leq \frac{M}{N}\sup_{\tau\in\overline{\Theta}}|\tau' D_0'\sigma_{u\epsilon}| = o(1),
\end{aligned} \tag{A.35}$$

where $\sup_{\tau\in\overline{\Theta}}\tau' D_0'\sigma_{u\epsilon}$ is bounded by Assumption 7. The term $\hat{A}_m$ has the decomposition $\hat{A}_m - A_m = \hat{A}_{m,1} + \hat{A}_{m,2} + \cdots + \hat{A}_{m,9} + o(1)$, where the $o(1)$ term is uniform in $m \leq M$ and consists of (A.33), (A.34), and (A.35), and

$$\hat{A}_{m,1} = D_0'\left(\frac{f'P_m f}{N} - A_m\right)D_0,$$

$$A_{m,2} = D_0'\frac{u'P_m f}{N}D_0,$$

$$\hat{A}_{m,3} = D_0'\frac{f'P_m u}{N}D_0,$$

$$\hat{A}_{m,4} = D_0'\frac{u'P_m u - E(u'P_m u|z)}{N}D_0,$$

$$\hat{A}_{m,5} = e_1 \frac{\epsilon' P_m u - m\sigma'_{u\epsilon}}{N} D_0,$$

$$\hat{A}_{m,6} = D'_0 \frac{u' P_m \epsilon - m\sigma_{u\epsilon}}{N} e'_1,$$

$$\hat{A}_{m,7} = e_1 \frac{\epsilon' P_m f}{N} D_0,$$

$$\hat{A}_{m,8} = D'_0 \frac{f' P_m \epsilon}{N} e'_1,$$

$$\hat{A}_{m,9} = \frac{\epsilon' P_m \epsilon - \sigma_\epsilon^2 m}{N} e_1 e'_1.$$

For $\hat{A}_{m,1}$, define $\hat{\Gamma}_{zz,m} = Z'_m Z_m/N$, $\hat{\Gamma}_{fz,m} = f' Z_m/N$, $\Gamma_{zz,k} = E[Z_{k,i} Z'_{k,i}]$, and $\Gamma_{fz,k} = E[f_i Z'_{k,i}]$, and choose a sequence M_1, where $M_1 \to \infty$ such that $M_1^3/N \to 0$. It then follows for $m \leq M_1$ that

$$\begin{aligned} &E[\|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\|^2] \\ &\quad = N^{-2} \sum_{i,j=1}^{N} \operatorname{tr} E[(f_i Z'_{m,i} - \Gamma_{fz,m})(f_j Z'_{m,j} - \Gamma_{fz,m})'] \\ &\quad = N^{-2} \sum_{i=1}^{N} \operatorname{tr} E[(f_i Z'_{m,i} - \Gamma_{fz,m})(f_i Z'_{m,i} - \Gamma_{fz,m})'] \\ &\quad = O\left(\frac{m}{N}\right) = o(1) \end{aligned}$$

and

$$\begin{aligned} &E[\|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\|^2] \\ &\quad = N^{-2} \sum_{i=1}^{N} \operatorname{tr} E[(Z_{m,i} Z'_{m,i} - \Gamma_{zz,m})(Z_{m,i} Z'_{m,i} - \Gamma_{zz,m})'] = O\left(\frac{m^2}{N}\right). \end{aligned}$$

Using the Markov inequality, one obtains

$$\begin{aligned} \Pr\Big(\sup_{m \leq M_1} \|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\| \geq \varepsilon\Big) &\leq \frac{M_1}{\varepsilon^2} \sup_{m \leq M_1} E[\|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\|^2] \\ &= O(M_1^2/N) = o(1) \end{aligned}$$

as well as

$$\begin{aligned} \Pr\Big(\sup_{m \leq M_1} \|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\| \geq \varepsilon\Big) &\leq \frac{M_1}{\varepsilon^2} \sup_{m \leq M_1} E[\|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\|^2] \\ &= O(M_1^3/N) = o(1). \end{aligned} \tag{A.36}$$

Let $\|C\|_2^2 = \sup \ell' C' C \ell / \ell' \ell$ for any matrix C, and note that $\|C_1 C_2\| \leq \|C_1\| \|C_2\|_2$ and $\|C_1 C_2\| \leq \|C_2\| \|C_1\|_2$ for any conforming matrices C_1 and C_2. Now,

$$
\begin{aligned}
\|\hat{\Gamma}_{fz,m} \hat{\Gamma}_{zz,m}^{-1} \hat{\Gamma}'_{fz,m} - A_m\| &\leq \|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\| \|\hat{\Gamma}_{zz,m}^{-1} \hat{\Gamma}'_{fz,m}\|_2 \\
&\quad + \|\Gamma_{fz,m}\|_2 \|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\| \|\hat{\Gamma}_{fz,m}\|_2 \\
&\quad + \|\Gamma_{fz,m} \Gamma_{zz,m}^{-1}\|_2 \|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\|, \\
\|\hat{\Gamma}_{zz,m}^{-1} \hat{\Gamma}'_{fz,m}\|_2 &\leq \|\hat{\Gamma}_{zz,m}^{-1} \hat{\Gamma}'_{fz,m} - \Gamma_{zz,m}^{-1} \Gamma'_{fz,m}\| + \|\Gamma_{zz,m}^{-1} \Gamma'_{fz,m}\|_2 \\
&\leq \|\hat{\Gamma}_{fz,m} - \Gamma_{fz,m}\| \|\hat{\Gamma}_{zz,m}^{-1}\|_2 + \|\Gamma_{fz,m}\|_2 \|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\| \\
&\quad + \|\Gamma_{zz,m}^{-1} \Gamma'_{fz,m}\|_2,
\end{aligned}
$$

and

$$
\|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\| \leq \|\hat{\Gamma}_{zz,m}^{-1}\|_2 \|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\| \|\Gamma_{zz,m}^{-1}\|_2.
$$

Define F such that $\|\Gamma_{zz,m}^{-1}\|_2 \leq F$, where F is finite by Assumption 6, and let

$$
\zeta_{m,N} := \|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\|_2 / (F \|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\|_2 + F^2) \leq \|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\|
$$

such that $\sup_{m \leq M_1} \zeta_{m,N} \leq \sup_{m \leq M_1} \|\hat{\Gamma}_{zz,m} - \Gamma_{zz,m}\| = o_p(1)$ by (A.36). Following Lewis and Reinsel (1985, p. 397),

$$
\begin{aligned}
\|\hat{\Gamma}_{zz,m}^{-1}\|_2 &\leq \|\Gamma_{zz,m}^{-1}\|_2 + \|\hat{\Gamma}_{zz,m}^{-1} - \Gamma_{zz,m}^{-1}\|_2 \\
&\leq F + F^2 \zeta_{m,N} / (1 - F \zeta_{m,N}) \\
&\leq F + F^2 \Big(\sup_{m \leq M_1} \zeta_{m,N} \Big) \Big/ \Big(1 - F \sup_{m \leq M_1} \zeta_{m,N} \Big) = O_p(1).
\end{aligned}
$$

It now follows that

$$
\begin{aligned}
\text{(A.37)} \qquad \sup_{m \leq M_1} \left\| \frac{f' P_m f}{N} - A_m \right\| &= \sup_{m \leq M_1} \|\hat{\Gamma}_{fz,m} \hat{\Gamma}_{zz,m}^{-1} \hat{\Gamma}'_{fz,m} - A_m\| \\
&= o_p(1).
\end{aligned}
$$

For $M_1 \leq m \leq M$, it follows that $A_m \to \bar{H} = E[f_i f_i']$. Then

$$
\begin{aligned}
\text{(A.38)} \qquad \frac{f' P_m f}{N} - A_m &= -f'(I - P_m)' f / N + f' f / N - \bar{H} + \bar{H} - A_m \\
&= o_p(1),
\end{aligned}
$$

where the $o_p(1)$ term is uniform in $M_1 \le m \le M$ because

$$\begin{aligned}
&\sup_{\tau\in\overline{\Theta}, M_1\le m\le M} \tau' f'(I-P_m)f\tau/N \\
&\quad \le \sup_{M_1\le m\le M} \frac{m^{2\alpha}}{M_1^{\alpha}}\Big(\sup_{\tau} \tau' f'(I-P_m)f\tau/N\Big) \\
&\quad = M_1^{-\alpha} O_p(1) = o_p(1)
\end{aligned}$$

by Assumption 2. By Assumption 6 and for $M_1 \le m \le M$, $\bar{H} - A_m = O(m^{-2\alpha}) \le O(M_1^{-2\alpha}) = o(1)$. By a law of large numbers,

$$f'f/N - \bar{H} = O_p(1/\sqrt{N}) = o_p(1).$$

Together, (A.37) and (A.38) imply that

$$\sup_{\tau\in\overline{\Theta}, m\le M} \|\hat{A}_{1,m}\| = o_p(1).$$

Now consider, for some $\varepsilon > 0$ not necessarily the same as in (A.32),

$$\begin{aligned}
\text{(A.39)}\quad &\Pr\Big(\sup_{\tau\in\overline{\Theta}, m\le M} |\tau'(\hat{A}_{m,2} + \cdots + \hat{A}_{m,9})\tau| > \varepsilon \Big| z\Big) \\
&\quad \le \sum_{j=2}^{9} \Pr\left(\sup_{\tau\in\overline{\Theta}} \|\tau\| \sum_{m=1}^{M} \|\hat{A}_{m,j}\| > \varepsilon \middle| z\right) \\
&\quad \le \frac{M \sup_{\tau} \|\tau\|}{\varepsilon} \max_{m\le M} \sum_{j=2}^{9} \sqrt{E[\|\hat{A}_{m,j}\|^2 | z]}.
\end{aligned}$$

To show that $M \max_{m\le M} E[(\epsilon' P_m \epsilon - \sigma_\epsilon^2 m)^2/N^2 | z] \to_p 0$, we observe that

$$\begin{aligned}
E[(\epsilon' P_m \epsilon - \sigma_\epsilon^2 m)^2 | z] &= \sigma_\epsilon^4 (\operatorname{tr} P_m)^2 + 2\sigma_\epsilon^4 (\operatorname{tr} P_m) - \sigma_\epsilon^4 m^2 \\
&\quad + \operatorname{Cum}[\epsilon_i, \epsilon_i, \epsilon_i, \epsilon_i] \sum_{i=1}^{N} (P_{m,ii})^2 \\
&= 2\sigma_\epsilon^4 m + \operatorname{Cum}[\epsilon_i, \epsilon_i, \epsilon_i, \epsilon_i] \sum_{i=1}^{N} (P_{m,ii})^2 \\
&= O(m) + o_p(m)
\end{aligned}$$

because $\sum_{i=1}^{N}(P_{m,ii})^2 \leq (\max_i P_{m,ii})\sum_{i=1}^{N} P_{m,ii} = o_p(m)$ by the same calculation as the proofs of Lemmas A.5(ii) and A.6(iv). Therefore,

$$
\begin{aligned}
\text{(A.40)}\quad & M \max_{m\leq M} E[(\epsilon' P_m \epsilon - \sigma_\epsilon^2 m)^2/N^2|z] \\
&\leq M \max_{m\leq M} \frac{2\sigma_\epsilon^4 m + \mathrm{Cum}[\epsilon_i,\epsilon_i,\epsilon_i,\epsilon_i]\sum_{i=1}^{N}(P_{m,ii})^2}{N^2} \\
&\leq \frac{M \max_{m\leq M}(\max_i P_{m,ii})m}{N^2} + \frac{2\sigma_\epsilon^4 M^2}{N^2} \\
&= O_p\left(\frac{M^2}{N^2}\right) + \frac{2\sigma_\epsilon^4 M^2}{N^2} \to_p 0.
\end{aligned}
$$

Similarly, we can show that $M\max_{m\leq M} E[\|\hat{A}_{m,4}\|^2|z] \to_p 0$, $M\max_{m\leq M} E[\|\hat{A}_{m,5}\|^2|z] \to_p 0$, and $M\max_{m\leq M} E[\|\hat{A}_{m,6}\|^2|z] \to_p 0$. Next,

$$
\begin{aligned}
& M \max_{m\leq M} E[\|D_0' f' P_m u D_0/N\|^2|z] \\
&\leq \|D_0\|^4 M \max_{m\leq M} E[\|f' P_m u/N\|^2|z] \\
&= \|D_0\|^4 M \max_{m\leq M} \frac{\mathrm{tr}(f' P_m E[uu'|z] P_m f)}{N^2} \\
&= \|D_0\|^4 \,\mathrm{tr}(\Sigma_u) M \max_{m\leq M} \frac{\mathrm{tr}(f' P_m f)}{N^2} \\
&\leq \|D_0\|^4 \,\mathrm{tr}(\Sigma_u) M \frac{\mathrm{tr}(f' f)}{N^2} \\
&= O_p\left(\frac{M}{N}\right) = o_p(1).
\end{aligned}
$$

Analogous calculations show that $M\max_{m\leq M} E[\|\hat{A}_{m,7}\|^2|z] = o(1)$ and $M \times \max_{m\leq M} E[\|\hat{A}_{m,8}\|^2|z] = o(1)$. Summing up, we have $\frac{M}{\varepsilon}\sup_{\tau\in\overline{\Theta}}\|\tau\| \times \sum_{j=1}^{9}\max_{m\leq M} E[\|\hat{A}_{m,j}\|^2|z] \to_p 0$. Combining (A.31), (A.32), (A.33), (A.34), (A.35), and (A.39) establishes that

$$
\text{(A.41)}\quad \sup_{\tau\in\overline{\Theta}, m\leq M} |L_{n,m}(\tau) - L_m(\tau)| = o_p(1).
$$

From (A.32) and the fact that if $\|\tau - \tau_0\| \geq \varepsilon$ for some $\varepsilon > 0$, there exists an $\eta > 0$ such that $\sup_{m\leq M} |L_m(\tau) - L_m(\tau_0)| \geq \eta$, it follows that

$$\Pr\Big(\sup_{m\leq M} |\hat{\beta}_{L,m} - \beta_0| \geq \varepsilon \big| z\Big) \leq \Pr\Big(\sup_{m\leq M} |L_m(\hat{\tau}_m) - L_m(\tau_0)| \geq \eta \big| z\Big)$$

with $\hat{\tau}_m = (1, -\hat{\beta}'_{L,m})'$ and by standard arguments,

$$\begin{aligned}|L_m(\hat{\tau}_m) - L_m(\tau_0)| \leq |L_{n,m}(\hat{\tau}_m) - L_m(\hat{\tau}_m)| + |L_{n,m}(\tau_0) - L_m(\tau_0)| \\ + |L_{n,m}(\hat{\tau}_m) - L_{n,m}(\tau_0)|,\end{aligned}$$

where $0 \leq L_{n,m}(\hat{\tau}_m) \leq L_{n,m}(\tau_0) + o_p(1) = o_p(1)$ uniformly in $m \leq M$ by the definition of $\hat{\tau}_m$ and Lemma A.11 such that

$$\begin{aligned}\sup_{m\leq M} |L_{n,m}(\hat{\tau}_m) - L_{n,m}(\tau_0)| &\leq 2 \sup_{m\leq M} |L_{n,m}(\tau_0)| + o_p(1) \\ &= o_p(1)\end{aligned}$$

and

$$\begin{aligned}\text{(A.42)}\quad &\sup_{m\leq M} |L_m(\hat{\tau}_m) - L_m(\tau_0)| \\ &\quad \leq 2 \sup_{\tau\in\overline{\Theta}, m\leq M} |L_{n,m}(\tau) - L_m(\tau)| + o_p(1) \\ &\quad = o_p(1)\end{aligned}$$

by (A.41). *Q.E.D.*

LEMMA A.13: *If Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*, *it follows that for* $\hat{\beta}$ *defined in* (2.3) (*MALIML*), $|\hat{\beta} - \beta_0| \rightarrow_p 0$.

PROOF: Let $A_m(\beta) \equiv (y - X\beta)' P_m (y - X\beta)/N$ and $B(\beta) \equiv (y - X\beta)'(y - X\beta)/N$. Define $\Lambda_m(\beta) \equiv A_m(\beta)/B(\beta)$.

As $\sup_{m\leq M} \|\hat{\beta}_{L,m} - \beta_0\| \rightarrow_p 0$ by Lemma A.12, it follows that $\sup_{m\leq M} |B(\hat{\beta}_{L,m}) - \sigma^2_\epsilon| \rightarrow_p 0$. Moreover,

$$\text{(A.43)}\quad A_m(\beta_0) = \epsilon' P_m \epsilon/N \rightarrow_p 0$$

uniformly in $m \leq M$ by Lemma A.11(i), which implies that $\sup_{m\leq M} |A_m(\hat{\beta}_{L,m})| \rightarrow_p 0$ and, therefore, $\sup_{m\leq M} |\Lambda_m(\hat{\beta}_{L,m})| \rightarrow_p 0$. We also note that $\Lambda_m(\beta) = L_{n,m}(\tau) \leq 1$ uniformly in $m \leq N$ and β.

It now follows that for $\Lambda(W)=\sum_{m=1}^{M} w_m\Lambda_m(\beta_0)$ that

$$\begin{aligned}|\hat{\Lambda}(W)-\Lambda(W)| &\leq \sum_{m=1}^{M}|w_m||L_{n,m}(\hat{\tau}_m)-L_{n,m}(\tau_0)| \\ &\leq 2\sup_{m,\tau}|L_{n,m}(\tau)-L_m(\tau)|\sum_{m=1}^{M}|w_m| \\ &\quad+\sup_{m\leq M}|L_m(\hat{\tau}_m)-L_m(\tau_0)|\sum_{m=1}^{M}|w_m|,\end{aligned}$$

where $2\sup_{m,\tau}|L_{n,m}(\tau)-L_m(\tau)|=o_p(1)$ by Lemma A.12, $\sup_{m\leq M}|L_m(\hat{\tau}_m)-L_m(\tau_0)|=o_p(1)$ by (A.42), and $\sum_{m=1}^{M}|w_m|=O(1)$. It now follows that

$$\hat{\beta}-\beta_0=(X'P(W)X-\hat{\Lambda}(W)X'X)^{-1}(X'P(W)\epsilon-\hat{\Lambda}(W)X'\epsilon). \tag{A.44}$$

We have $(\hat{\Lambda}(W)-\Lambda(W))X'X/N=o_p(1)$ and $|\Lambda(W)|\leq\sum_{m=1}^{M}|w_m||\Lambda_m(\beta_0)|=o_p(1)$ such that

$$N^{-1}(X'P(W)X-\hat{\Lambda}(W)X'X)=N^{-1}X'P(W)X+o_p(1) \tag{A.45}$$

and, similarly, $\hat{\Lambda}(W)X'\epsilon/N=o_p(1)$ such that

$$\hat{\beta}-\beta_0=(X'P(W)X)^{-1}X'P(W)\epsilon+o_p(1),$$

and the result follows from Theorem A.1. *Q.E.D.*

LEMMA A.14: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*. *Then*, *for* $\hat{\beta}$ *defined in* (2.3) (*MALIML*), $\sqrt{N}(\hat{\beta}-\beta_0)\to_d N(0,\sigma_\epsilon^2\bar{H}^{-1})$.

PROOF: The result follows from (A.44), (A.45), and the fact that $X'\epsilon/\sqrt{N}=O_p(1)$ together with $\hat{\Lambda}(W)=o_p(1)$. We then have

$$\sqrt{N}(\hat{\beta}-\beta_0)=\left(\frac{X'P(W)X}{N}\right)^{-1}\frac{X'P(W)\epsilon}{\sqrt{N}}+o_p(1)$$

such that the result again follows from Theorem A.1. *Q.E.D.*

LEMMA A.15: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*. *Let* $\Lambda_{\beta\beta,m}(\beta)$ *be the Hessian of* $\Lambda_m(\beta)$. *If* $\sup_{m\leq M}\|\tilde{\beta}_m-\beta_0\|\to_p 0$, *then*

$$\sup_{m\leq M}\|\Lambda_{\beta\beta,m}(\beta_0)-\Lambda_{\beta\beta,m}(\tilde{\beta}_m)\|=o_p(1)$$

and

$$\sup_{m\leq M}\left\|\Lambda_{\beta\beta,m}(\beta_0)-\frac{2}{\sigma_\epsilon^2}\bar{H}_m\right\|=o_p(1).$$

PROOF: Let $\Lambda_{\beta,m}(\beta)$ and $\Lambda_{\beta\beta,m}(\beta)$ be the gradient and the Hessian of $\Lambda_m(\beta)$. Let $A_m(\beta)\equiv(y-X\beta)'P_m(y-X\beta)/N$ and $B(\beta)=(y-X\beta)'(y-X\beta)/N$. Let $A_{\beta,m}(\beta)$ and $B_\beta(\beta)$ be the gradients of $A_m(\beta)$ and $B(\beta)$, respectively, and let $A_{\beta\beta,m}(\beta)$ and $B_{\beta\beta}(\beta)$ be the Hessians of $A_m(\beta)$ and $B(\beta)$, respectively. We have

$$\begin{aligned}\Lambda_{\beta,m}(\beta)&=B(\beta)^{-1}(A_{\beta,m}(\beta)-\Lambda_m(\beta)B_\beta(\beta)),\\ \Lambda_{\beta\beta,m}(\beta)&=B(\beta)^{-1}(A_{\beta\beta,m}(\beta)-\Lambda_m(\beta)B_{\beta\beta}(\beta))\\ &\qquad-B(\beta)^{-1}(B_\beta(\beta)\Lambda_{\beta,m}(\beta)'+\Lambda_{\beta,m}(\beta)B_\beta(\beta)').\end{aligned}$$

By assumption, $\sup_{m\leq M}\|\tilde{\beta}_m-\beta_0\|\to_p 0$, which implies that $\sup_{m\leq M}|B(\tilde{\beta}_m)-\sigma_\epsilon^2|\to_p 0$ and $\sup_{m\leq M}|B_\beta(\tilde{\beta}_m)-(-2\sigma_{u\epsilon})|\to_p 0$. Moreover,

$$\max_{m\leq M}|A_m(\beta_0)|=\max_{m\leq M}|\epsilon'P_m\epsilon/N|\to_p 0$$

by Lemma A.11(i) and

$$\begin{aligned}\text{(A.46)}\quad \max_{m\leq M}\|A_{\beta,m}(\beta_0)\|&=\max_{m\leq M}\|X'P_m\epsilon/N\|\\ &\leq\max_{m\leq M}\|f'P_m\epsilon/N\|+\max_{m\leq M}\|u'P_m\epsilon/N\|=o_p(1),\end{aligned}$$

where $\max_{m\leq M}\|f'P_m\epsilon/N\|=o_p(1)$ by Lemma A.11(ii) and $\max_{m\leq M}\|u'P_m\epsilon/N\|=o_p(1)$ by Lemma A.11(iii).

From the proof of Lemma A.12 and (A.41), it follows that $\sup_{m\leq M}\Lambda_m(\tilde{\beta}_m)\to_p 0$. Similarly, we note that

$$\begin{aligned}A_{\beta,m}(\tilde{\beta}_m)&=X'P_m(y-X\tilde{\beta}_m)/N\\ &=X'P_m\epsilon/N+X'P_mX(\tilde{\beta}_m-\beta_0)/N,\end{aligned}$$

where $\epsilon'P_mX/N=o_p(1)$ uniformly in $m\leq M$ by (A.46) and $X'P_mX/N$ is uniformly bounded by the same arguments as in the proof of Lemma A.12. This shows that $A_{\beta,m}(\tilde{\beta}_m)\to_p 0$ and, therefore, $\Lambda_{\beta,m}(\tilde{\beta}_m)\to_p 0$. Now, consider

$$\begin{aligned}&\Lambda_{\beta\beta,m}(\tilde{\beta}_m)-\Lambda_{\beta\beta,m}(\beta_0)\\ &\quad=\left(\frac{1}{\epsilon'\epsilon/N}-B(\tilde{\beta}_m)^{-1}\right)2\left(\frac{X'P_mX}{N}-\Lambda_m(\beta_0)\frac{X'X}{N}\right)\end{aligned}$$

$$
\begin{aligned}
&-\left(\frac{1}{\epsilon'\epsilon/N}-B(\tilde{\beta}_m)^{-1}\right)\left(\frac{\epsilon'X}{N}\Lambda_{\beta,m}(\beta_0)'+\Lambda_{\beta,m}(\beta_0)\frac{X'\epsilon}{N}\right)\\
&+B(\tilde{\beta}_m)^{-1}(\Lambda_m(\beta_0)-\Lambda_m(\tilde{\beta}_m))X'X/N\\
&-B(\tilde{\beta})^{-1}(B_\beta(\tilde{\beta}_m)\Lambda_{\beta,m}(\tilde{\beta}_m)'-B_\beta(\beta_0)\Lambda_{\beta,m}(\beta_0)')\\
&-B(\tilde{\beta}_m)^{-1}(\Lambda_{\beta,m}(\tilde{\beta}_m)B_\beta(\tilde{\beta}_m)'-\Lambda_{\beta,m}(\beta_0)B_\beta(\beta_0)'),
\end{aligned}
$$

where

$$
\begin{aligned}
&\left(\frac{1}{\epsilon'\epsilon/N}-B(\tilde{\beta}_m)^{-1}\right)\\
&\quad=\frac{B(\tilde{\beta}_m)-\epsilon'\epsilon/N}{B(\tilde{\beta}_m)\epsilon'\epsilon/N}\\
&\quad=\frac{2\epsilon'X/N(\tilde{\beta}_m-\beta_0)+(\tilde{\beta}-\beta_0)'X'X(\tilde{\beta}_m-\beta_0)/N}{B(\tilde{\beta}_m)\epsilon'\epsilon/N}\\
&\quad=o_p(1)
\end{aligned}
$$

uniformly in $m\le M$. Since $X'P_mX/N-\Lambda_m(\beta_0)X'X/N=O_p(1)$ uniformly in $m\le M$ and all other terms are of smaller order, it follows that $\sup_{m\le M}\|\Lambda_{\beta\beta,m}(\beta_0)-\Lambda_{\beta\beta,m}(\tilde{\beta}_m)\|=o_p(1)$. Next consider

$$
\begin{aligned}
&\Lambda_{\beta\beta,m}(\beta_0)-\frac{2}{\sigma_\epsilon^2}\bar{H}_m\\
&\quad=\left(\frac{1}{\epsilon'\epsilon/N}-\frac{1}{\sigma_\epsilon^2}\right)\frac{2X'P_mX}{N}+\frac{2}{\sigma_\epsilon^2}\left(\frac{X'P_mX}{N}-\bar{H}_m\right)\\
&\qquad-\frac{1}{\epsilon'\epsilon/N}\left(\Lambda_m(\beta_0)\frac{X'X}{N}+2\frac{\epsilon'X}{N}\Lambda_{\beta,m}(\beta_0)'+2\Lambda_{\beta,m}(\beta_0)\frac{X'\epsilon}{N}\right).
\end{aligned}
$$

Note that $2X'P_mX/N-2\bar{H}_m\to_p 0$, where the convergence is uniform in $m\le M$ by the same arguments as in the proof of Lemma A.12. Also note that $B_{\beta\beta}(\beta)=2X'X/N\to_p 2E(X_iX_i')$. It therefore follows that $\sup_{m\le M}\|\Lambda_{\beta\beta,m}(\beta_0)-\frac{2}{\sigma_\epsilon^2}\bar{H}_m\|=o_p(1)$ uniformly in $m\le M$. *Q.E.D.*

LEMMA A.16: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*. *Then*

$$
\begin{aligned}
\sqrt{N}(\hat{\beta}_{L,m}-\beta_0)=(\bar{H}_m^{-1}+o_p(1))&\left(h-\frac{f'(I-P_m)\epsilon}{\sqrt{N}}+\frac{v'P_m\epsilon}{\sqrt{N}}\right)\\
&+o_p(1),
\end{aligned}
$$

where both $o_p(1)$ terms are uniform in $m\le M$.

PROOF: Let $\Lambda_{\beta,m}(\beta)$ and $\Lambda_{\beta\beta,m}(\beta)$ be the gradient and the Hessian of $\Lambda_m(\beta)$, respectively. A standard Taylor expansion shows that

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{L,m}-\beta_0) &= -\Lambda_{\beta\beta,m}(\tilde{\beta}_m)^{-1}\sqrt{N}\Lambda_{\beta,m}(\beta_0) \\ &= \left(\frac{\tilde{\sigma}_\epsilon^2\Lambda_{\beta\beta,m}(\tilde{\beta}_m)}{2}\right)^{-1}\left(-\frac{\tilde{\sigma}_\epsilon^2\sqrt{N}\Lambda_{\beta,m}(\beta_0)}{2}\right)\end{aligned}$$

for some mean value $\tilde{\beta}_m$, where $\tilde{\sigma}_\epsilon^2=\epsilon'\epsilon/N$. As $\sup\|\hat{\beta}_{L,m}-\beta_0\|=o_p(1)$ by Lemma A.12, it follows that $\sup_m\|\tilde{\beta}_m-\beta_0\|\to_p 0$ such that, by Lemma A.15, it follows that

$$\sqrt{N}(\hat{\beta}_{L,m}-\beta_0)=(\bar{H}_m^{-1}+o_p(1))\left(-\frac{\tilde{\sigma}_\epsilon^2\sqrt{N}\Lambda_{\beta,m}(\beta_0)}{2}\right),$$

where the $o_p(1)$ term is uniform in $m\le M$.

Consider the gradient term. Define $\hat{\alpha}=X'\epsilon/\epsilon'\epsilon$, $\alpha=\sigma_{u\epsilon}/\sigma_\epsilon^2$, and $v=u-\epsilon\alpha'$. It holds that $\hat{\alpha}-\alpha=O_p(1/\sqrt{N})$ by the CLT. We have the decomposition

$$\begin{aligned}-\frac{\tilde{\sigma}_\epsilon^2\sqrt{N}\Lambda_{\beta,m}(\beta_0)}{2} &= \frac{X'P_m\epsilon}{\sqrt{N}}-\frac{\epsilon'P_m\epsilon X'\epsilon}{\sqrt{N}\epsilon'\epsilon} \\ &= h-\frac{f'(I-P_m)\epsilon}{\sqrt{N}}+\frac{v'P_m\epsilon}{\sqrt{N}}-\sqrt{N}(\hat{\alpha}-\alpha)\frac{\epsilon'P_m\epsilon}{N}.\end{aligned}$$

First, we have $h\to_d N(0,\sigma^2\bar{H})$ by the CLT. Lemma A.11(ii) implies that $f'(I-P_m)\epsilon/\sqrt{N}=O_p(1)$ uniformly in $m\le M$. From Lemma A.11(i) $\sup_{m\le M}\epsilon'P_m\epsilon/N=o_p(1)$ such that $\sqrt{N}(\hat{\alpha}-\alpha)\epsilon'P_m\epsilon/N=o_p(1)$ uniformly in $m\le M$. In conclusion, we have

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{L,m}-\beta_0) &= (\bar{H}_m^{-1}+o_p(1))\left(h-\frac{f'(I-P_m)\epsilon}{\sqrt{N}}+\frac{v'P_m\epsilon}{\sqrt{N}}\right) \\ &\quad+o_p(1),\end{aligned}$$

where both $o_p(1)$ terms are small uniformly in $m\le M$. *Q.E.D.*

LEMMA A.17: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*. *Then*

$$\begin{aligned}\hat{\Lambda}(W) &= \tilde{\Lambda}(W)-\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2}-1\right)\tilde{\Lambda}(W)-\Lambda_q(W)+\hat{R}_\Lambda \\ &= \tilde{\Lambda}(W)+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{K'W}{N^{3/2}}\right)+o_p\left(\frac{\rho_{W,N}}{\sqrt{N}}\right),\end{aligned}$$

where

$$\Lambda_q(W) = \frac{1}{2}\sum_{m=1}^{M} w_m\big(\Lambda_{\beta,m}(\beta_0)'(\Lambda_{\beta\beta,m}(\beta_0))^{-1}\Lambda_{\beta,m}(\beta_0)\big)$$
$$= O_p\left(\frac{1}{N}\right),$$
$$\tilde{\Lambda}(W) = \frac{\epsilon' P(W)\epsilon}{N\sigma_\epsilon^2}$$
$$= \frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right),$$

$\tilde{\sigma}_\epsilon^2 = \epsilon'\epsilon/N$, $\sqrt{N}\hat{R}_\Lambda = O_p(1/\sqrt{N})$, $\hat{R}_\Lambda$ *is simply the difference between* $\hat{\Lambda}$ *and the first three terms in the expression between two equalities*, $\Lambda_{\beta,m}(\beta)$ *and* $\Lambda_{\beta\beta,m}(\beta)$ *are the gradient and the Hessian of* $\Lambda_m(\beta)$, *and* $\rho_{W,N} = \operatorname{tr}(S(W))$ *for* $S(W)$ *defined in* (A.6).

PROOF: We note that, in the LIML case, to show $o_p(\rho_{W,N})$, it is enough to show $o_p(W'\Gamma W/N + K'W/N + \sum_i (P_{ii}(W))^2/N + \Delta(W))$. We use the notation developed in the proof of Lemma A.14. We expand $\hat{\Lambda}_m = \Lambda_m(\hat{\beta}_{L,m})$ around the true value β_0. By Donald and Newey (2001, p. 1186),

$$\Lambda_m(\beta_0) = \tilde{\Lambda}_m - \left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}_m + \frac{(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2)^2}{\tilde{\sigma}_\epsilon^2\sigma_\epsilon^2}\tilde{\Lambda}_m,$$

where $\tilde{\Lambda}_m = \epsilon' P_m \epsilon/(N\sigma_\epsilon^2)$ such that

$$\sum_{m=1}^{M} w_m \Lambda_m(\beta_0) = \tilde{\Lambda}(W) - \left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}(W)$$
$$+ \frac{(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2)^2}{\tilde{\sigma}_\epsilon^2\sigma_\epsilon^2}\tilde{\Lambda}(W).$$

By a similar argument as in Lemma A.6(iv), we have

$$\tilde{\Lambda}(W) = \frac{\epsilon' P(W)\epsilon}{N\sigma_\epsilon^2} \tag{A.47}$$
$$= \frac{K'W}{N} + \frac{\epsilon' P(W)\epsilon - \sigma_\epsilon^2 K'W}{N\sigma_\epsilon^2}$$

$$= \frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right).$$

Consider

$$\begin{aligned}
\frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\beta)]'}{\partial\beta} &= -B(\beta)^{-2}B_\beta(\beta)\operatorname{vec}\big[(A_{\beta\beta,m}(\beta) - \Lambda_m(\beta)B_{\beta\beta}(\beta)) \\
&\quad - (B_\beta(\beta)\Lambda_{\beta,m}(\beta)' + \Lambda_{\beta,m}(\beta)B_\beta(\beta)')\big]' \\
&\quad - B(\beta)^{-1}\Lambda_{\beta,m}(\beta)\operatorname{vec}\big[B_{\beta\beta}(\beta))\big]' \\
&\quad - B_{\beta\beta}(\beta)\big[(K_{1,n}\otimes I)(\Lambda_{\beta,m}(\beta)\otimes I)\big]' \\
&\quad - \Lambda_{\beta\beta,m}(\beta)\big[(K_{1,n}\otimes I)(I\otimes B_\beta(\beta))\big]' \\
&\quad - B_{\beta\beta}(\beta)\big[(K_{1,n}\otimes I)(I\otimes \Lambda_{\beta,m}(\beta))\big]' \\
&\quad - \Lambda_{\beta\beta,m}(\beta)\big[(K_{1,n}\otimes I)(B_\beta(\beta)\otimes I)\big]',
\end{aligned}$$

where the result follows from Magnus and Neudecker (1988, p. 185) and $K_{1,n}$ is the commutation matrix. Let $\tilde{\beta}_m$ be some mean value between $\hat{\beta}_{L,m}$ and β_0. Then

$$\frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\tilde{\beta}_m)]'}{\partial\beta} - \frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\beta_0)]'}{\partial\beta} = o_p(1)$$

uniformly in $m \le M$ and $\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\beta_0)]'/\partial\beta$ is bounded uniformly in probability over $m \le M$. A Taylor expansion then leads to

$$\begin{aligned}
\sum_{m=1}^M w_m\hat{\Lambda}_m &= \sum_{m=1}^M w_m\Lambda_m(\beta_0) \\
&\quad - \sum_{m=1}^M w_m\frac{1}{2}(\hat{\beta}_{L,m}-\beta_0)'\Lambda_{\beta\beta,m}(\beta_0)(\hat{\beta}_{L,m}-\beta_0) \\
&\quad + \sum_{m=1}^M w_m(\hat{\beta}_{L,m}-\beta_0)'\frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\tilde{\beta}_m)]'}{\partial\beta} \\
&\quad \times ((\hat{\beta}_{L,m}-\beta_0)\otimes(\hat{\beta}_{L,m}-\beta_0)') \\
&= \sum_{m=1}^M w_m\Lambda_m(\beta_0)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}\sum_{m=1}^{M} w_m \Lambda_{\beta,m}(\beta_0)'(\Lambda_{\beta\beta,m}(\beta_0))^{-1}\Lambda_{\beta,m}(\beta_0) \\
&+O_p\left(\frac{1}{N^{3/2}}\right),
\end{aligned}
$$

where $O_p(1/N^{3/2})$ can be established by considering

$$
\begin{aligned}
&\left\|\sum_{m=1}^{M} w_m(\hat{\beta}_{L,m}-\beta_0)'\frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\tilde{\beta}_m)]'}{\partial\beta}\right. \\
&\qquad \left.\times\left((\hat{\beta}_{L,m}-\beta_0)\otimes(\hat{\beta}_{L,m}-\beta_0)'\right)\right\| \\
&\quad \le \sup_{m\le M}\left\|\frac{\partial \operatorname{vec}[\Lambda_{\beta\beta,m}(\tilde{\beta}_m)]'}{\partial\beta}\right\|\sum_{m=1}^{M}|w_m|\|\hat{\beta}_{L,m}-\beta_0\|^3
\end{aligned}
$$

with

$$
\begin{aligned}
&\sqrt{N}(\hat{\beta}_{L,m}-\beta_0) \\
&\quad =(\bar{H}_m^{-1}+o_p(1))\left(h-\frac{f'(I-P_m)\epsilon}{\sqrt{N}}+\frac{v'P_m\epsilon}{\sqrt{N}}\right)+o_p(1) \\
&\quad =O_p(1)+(\bar{H}_m^{-1}+o_p(1))\frac{v'P_m\epsilon}{\sqrt{N}},
\end{aligned}
$$

where the $O_p(1)$ and $o_p(1)$ terms are uniform in $m\le M$ such that

$$
\begin{aligned}
&\sum_{m=1}^{M}|w_m|\|\hat{\beta}_{L,m}-\beta_0\|^3 \\
&\quad \le O_p\big(N^{-3/2}\big)O_p\left(1+\sum_{m=1}^{M}|w_m|(\|\bar{H}_m^{-1}\|^3\|v'P_m\epsilon/\sqrt{N}\|^3\right. \\
&\qquad \left.+\|\bar{H}_m^{-1}\|^2\|v'P_m\epsilon/\sqrt{N}\|^2)\right) \\
&\qquad +O_p\big(N^{-3/2}\big)O_p\left(\sum_{m=1}^{M}|w_m|\|\bar{H}_m^{-1}\|\|v'P_m\epsilon/\sqrt{N}\|\right) \\
&\qquad +o_p\left(N^{-3/2}\sum_{m=1}^{M}|w_m|(\|\bar{H}_m^{-1}\|^3\|v'P_m\epsilon/\sqrt{N}\|^3)\right).
\end{aligned}
$$

Consider

$$\begin{aligned}
&\sum_{m=1}^{M} |w_m| \|\bar{H}_m^{-1}\|^3 E[\|v' P_m \epsilon/\sqrt{N}\|^3 | z] \\
&\quad \leq \sum_{m=1}^{M} |w_m| \|\bar{H}_m^{-1}\|^3 \big(E[\|v' P_m \epsilon/\sqrt{N}\|^4 | z]\big)^{3/4}
\end{aligned}$$

with

$$\begin{aligned}
E[\|v' P_m \epsilon/\sqrt{N}\|^4 | z] &= N^{-2} E\big[(\operatorname{tr}(v' P_m \epsilon\epsilon' P_m v))^2 | z\big] \\
&= N^{-2} \sum_{j_1, j_2} \sum_{i_1, \ldots, i_8 = 1}^{N} E\big[v_{j_1, i_1} v_{j_1, i_4} v_{j_2, i_5} v_{j_2, i_8} \epsilon_{i_2} \epsilon_{i_3} \epsilon_{i_6} \epsilon_{i_7} | z\big] \\
&\quad \times P_{m, i_1 i_2} P_{m, i_3 i_4} P_{m, i_5 i_6} P_{m, i_7 i_8} \\
&\leq C N^{-2} \sum_{j_1, j_2} \Bigg| \sum_{i_1, \ldots, i_4 = 1}^{N} \big(P_{m, i_1 i_2} P_{m, i_2 i_1} P_{m, i_3 i_4} P_{m, i_4 i_3} \\
&\quad + 2 P_{m, i_1 i_2} P_{m, i_4 i_1} P_{m, i_3 i_4} P_{m, i_2 i_3}\big) \Bigg| \\
&\quad + C N^{-2} \sum_{j_1, j_2} \Bigg| \sum_{i_1, \ldots, i_4 = 1}^{N} \big(P_{m, i_1 i_2} P_{m, i_2 i_3} P_{m, i_1 i_4} P_{m, i_4 i_3} \\
&\quad + 2 P_{m, i_1 i_2} P_{m, i_4 i_3} P_{m, i_1 i_4} P_{m, i_2 i_3}\big) \Bigg| \\
&\quad + C N^{-2} \sum_{j_1, j_2} \Bigg| \sum_{i_1, \ldots, i_4 = 1}^{N} \big(P_{m, i_1 i_2} P_{m, i_2 i_3} P_{m, i_3 i_4} P_{m, i_4 i_1} \\
&\quad + 2 P_{m, i_1 i_2} P_{m, i_3 i_2} P_{m, i_3 i_4} P_{m, i_4 i_1}\big) \Bigg| \\
&\quad + \text{lower order terms},
\end{aligned}$$

where C is a constant such that $(E[|\epsilon_i|^8 | z])^{1/2} (\max_a E[|v_{a,i}|^8 | z])^{1/2} \leq C$ and we use the fact that P_m is idempotent and symmetric such that $P_{m, i_1 i_2} = P_{m, i_2 i_1}$ and

$\sum_{i_2=1}^{N} P_{m,i_1 i_2} P_{m,i_2 i_3} = P_{m,i_1 i_3}$. This implies, for example, that

$$\begin{aligned}
&\sum_{i_1,\ldots,i_4=1}^{N} P_{m,i_1 i_2} P_{m,i_4 i_1} P_{m,i_3 i_4} P_{m,i_2 i_3} \\
&\quad = \sum_{i_1,\ldots,i_3=1}^{N} P_{m,i_1 i_2} P_{m,i_2 i_3} \sum_{i_4=1}^{N} P_{m,i_1 i_4} P_{m,i_4 i_3} \\
&\quad = \sum_{i_1,i_3=1}^{N} P_{m,i_1 i_3} \sum_{i_2=1}^{N} P_{m,i_1 i_2} P_{m,i_2 i_3} \\
&\quad = \sum_{i_1,i_3=1}^{N} P_{m,i_1 i_3}^2 = \operatorname{tr}(P_m P_m) = m,
\end{aligned}$$

$\sum_{i_1,\ldots,i_4=1}^{N} P_{m,i_1 i_2} P_{m,i_2 i_3} P_{m,i_1 i_4} P_{m,i_4 i_3} = m$, and $\sum_{i_1,\ldots,i_4=1}^{N} P_{m,i_1 i_2} P_{m,i_4 i_3} P_{m,i_1 i_4} P_{m,i_2 i_3} = m$ and $\sum_{i_1,i_2=1}^{N} P_{m,i_1 i_2}^2 \sum_{i_3,i_4=1}^{N} P_{m,i_3 i_4}^2 = m^2$. This implies that

$$E[\|v' P_m \epsilon/\sqrt{N}\|^4 | z] = O(m^2/N^2) = o(1)$$

uniformly in $m \le M$ and, by the Markov inequality and the fact that $\|\bar{H}_m^{-1}\|$ is bounded uniformly in m, that

$$\begin{aligned}
&\sum_{m=1}^{M} |w_m| (\|\bar{H}_m^{-1}\|^3 \|v' P_m \epsilon/\sqrt{N}\|^3 + \|\bar{H}_m^{-1}\|^2 \|v' P_m \epsilon/\sqrt{N}\|^2 \\
&\qquad + \|\bar{H}_m^{-1}\| \|v' P_m \epsilon/\sqrt{N}\|) \\
&\quad = o_p(1).
\end{aligned}$$

Thus, we have shown that $\sum_{m=1}^{M} |w_m| \|\hat{\beta}_{L,m} - \beta_0\|^3 = O_p(N^{-3/2})$. To summarize, it then follows that

$$\hat{\Lambda}(W) = \sum_{m=1}^{M} w_m \Lambda_m(\beta_0) - \Lambda_q(W) + O_p\left(\frac{1}{N^{3/2}}\right).$$

Since $O_p(N^{-3/2}) = N^{-1/2} o_p(W' \Gamma W/N)$, it follows that $\sqrt{N} \hat{R}_\Lambda = o_p(\rho_{W,N})$. Now turn to $\Lambda_q(W)$, where by Lemma A.15,

$$\begin{aligned}
&\Lambda_{\beta,m}(\beta_0)' (\Lambda_{\beta\beta,m}(\beta_0))^{-1} \Lambda_{\beta,m}(\beta_0) \\
&\quad = \left(\frac{h}{\sqrt{N}} - \frac{f'(I - P_m)\epsilon}{N} + \frac{v' P_m \epsilon}{N} \right)' (\bar{H}_m^{-1} + o_p(1))
\end{aligned}$$

$$\times \left(\frac{h}{\sqrt{N}} - \frac{f'(I-P_m)\epsilon}{N} + \frac{v'P_m\epsilon}{N} \right) + o_p(1)$$

$$= \frac{h'\bar{H}_m^{-1}h}{N} - N^{-3/2}h'\bar{H}_m^{-1}f'(I-P_m)\epsilon + N^{-3/2}h'\bar{H}_m^{-1}v'P_m\epsilon$$

$$+ N^{-3/2}\epsilon'(I-P_m)f\bar{H}_m^{-1}h$$

$$+ N^{-2}\epsilon'(I-P_m)f\bar{H}_m^{-1}f'(I-P_m)\epsilon + N^{-2}\epsilon'(I-P_m)f\bar{H}_m^{-1}v'P_m\epsilon$$

$$+ N^{-3/2}\epsilon'P_m v\bar{H}_m^{-1}h + N^{-2}\epsilon'P_m v\bar{H}_m^{-1}f'(I-P_m)\epsilon$$

$$+ N^{-2}\epsilon'P_m v\bar{H}_m^{-1}v'P_m\epsilon$$

$$+ \text{terms of lower order.}$$

Next, consider

$$N^{-3/2}\left\| \sum_{m=1}^{M} w_m h'\bar{H}_m^{-1}f'(I-P_m)\epsilon \right\|$$

$$\leq \|h/N\| \sum_{m=1}^{M} |w_m| \|\bar{H}_m^{-1}\| \|f'(I-P_m)\epsilon/\sqrt{N}\|$$

$$= O_p(N^{-1}),$$

where $\sup_{m\leq M}\|\epsilon'(I-P_m)f/\sqrt{N}\| = O_p(1)$ by Lemma A.11(ii). For $N^{-3/2}h' \times \bar{H}_m^{-1}v'P_m\epsilon$, note that

$$E[\|v'P_m\epsilon/\sqrt{N}\|^2|z] = \operatorname{tr} E[v'P_m\epsilon\epsilon'P_m v/N|z] \tag{A.48}$$

$$= \frac{m\sigma_\epsilon^2}{N}\operatorname{tr}\Sigma_v + \frac{\operatorname{Cum}[v_i, v_i, \epsilon_i, \epsilon_i]}{N}\sum_{i=1}^{n}(P_{m,ii})^2$$

such that, by the Markov inequality,

$$N^{-3/2}\left\| \sum_{m=1}^{M} w_m h'\bar{H}_m^{-1}v'P_m\epsilon \right\|$$

$$\leq \|h/N\| \sum_{m=1}^{M} |w_m| \|\bar{H}_m^{-1}\| \|v'P_m\epsilon/\sqrt{N}\|$$

$$\leq O_p(N^{-1})O_p\left(\sum_{m=1}^{M} |w_m|\sqrt{m/N} \right) = O_p(N^{-1}).$$

For $N^{-2}\epsilon'(I-P_m)f\bar{H}_m^{-1}f'(I-P_m)\epsilon$, note that

$$\begin{aligned}
&N^{-2}\left\|\sum_{m=1}^{M} w_m\epsilon'(I-P_m)f\bar{H}_m^{-1}f'(I-P_m)\epsilon\right\| \\
&\quad\le N^{-1}\sum_{m=1}^{M}|w_m|\|\bar{H}_m^{-1}\|\|f'(I-P_m)\epsilon/\sqrt{N}\|^2 \\
&\quad= O_p(N^{-1}).
\end{aligned}$$

For $N^{-2}\epsilon'(I-P_m)f\bar{H}_m^{-1}v'P_m\epsilon$, note that

$$\begin{aligned}
&N^{-2}\left\|\sum_{m=1}^{M} w_m\epsilon'(I-P_m)f\bar{H}_m^{-1}v'P_m\epsilon\right\| \\
&\quad\le N^{-1}\sup_{m\le M}\|\epsilon'(I-P_m)f/\sqrt{N}\|\sum_{m=1}^{M}|w_m|\|v'P_m\epsilon/\sqrt{N}\| \\
&\quad= o_p(N^{-1})
\end{aligned}$$

by Lemma A.11 and (A.48). For $N^{-3/2}\epsilon'P_m v\bar{H}_m^{-1}h$, it follows that

$$\begin{aligned}
N^{-3/2}\left\|\sum_{m=1}^{M} w_m\epsilon'P_m v\bar{H}_m^{-1}h\right\| &\le \|h/N\|\sum_{m=1}^{M}|w_m|\|\bar{H}_m^{-1}\|\|v'P_m\epsilon/\sqrt{N}\| \\
&= o_p(N^{-1})
\end{aligned}$$

by (A.48) and the Markov inequality. For $N^{-2}\epsilon'P_m v\bar{H}_m^{-1}v'P_m\epsilon$, it holds that

$$\begin{aligned}
N^{-2}\left\|\sum_{m=1}^{M} w_m\epsilon'P_m v\bar{H}_m^{-1}v'P_m\epsilon\right\| &\le N^{-1}\sum_{m=1}^{M}|w_m|\|\bar{H}_m^{-1}\|\|\epsilon'P_m v/\sqrt{N}\|^2 \\
&= o_p(N^{-1})
\end{aligned}$$

by (A.48) and the Markov inequality. Together these results imply that

$$\begin{aligned}
\text{(A.49)}\quad \sum_{m=1}^{M} w_m\big(\Lambda_{\beta,m}(\beta_0)'(\Lambda_{\beta\beta,m}(\beta_0))^{-1}\Lambda_{\beta,m}(\beta_0)\big) &= \frac{h'\bar{H}^{-1}(W)h}{N} + O_p(N^{-1}) \\
&= O_p(N^{-1}),
\end{aligned}$$

where $\bar{H}^{-1}(W) = \sum_{m=1}^{M} w_m\bar{H}_m^{-1}$ and $\|\bar{H}^{-1}(W)\| \le \sum_{m=1}^{M}|w_m|\|\bar{H}_m^{-1}\| = O(1)$.

To sum up, we have

$$\begin{aligned}\hat{\Lambda}(W) &= \sum_{m=1}^{M} w_m \Lambda_m(\beta_0) - \Lambda_q(W) + O_p\left(\frac{1}{N^{3/2}}\right) \\ &= \tilde{\Lambda}(W) - \left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}(W) + \frac{(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2)^2}{\tilde{\sigma}_\epsilon^2 \sigma_\epsilon^2}\tilde{\Lambda}(W) \\ &\quad - \Lambda_q(W) + O_p\left(\frac{1}{N^{3/2}}\right) \\ &= \tilde{\Lambda}(W) - \left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}(W) - \Lambda_q(W) + O_p\left(\frac{1}{N^{3/2}}\right),\end{aligned}$$

where the last equality follows by $(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2)^2 = O_p(1/N)$. This proves the first equality in the lemma.

We now consider the second equality in the lemma. We have from (A.47) that

$$\begin{aligned}&\left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}(W) \\ &= O_p\left(\frac{1}{\sqrt{N}}\right)\left(\frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right)\right) \\ &= O_p\left(\frac{K'W}{N^{3/2}}\right) + o_p\left(\frac{\rho_{W,N}}{\sqrt{N}}\right).\end{aligned}$$

We also have that

$$\Lambda_q(W) = O_p\left(\frac{1}{N}\right)$$

from (A.49). It therefore follows that

$$\hat{\Lambda}(W) = \tilde{\Lambda}(W) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W}{N^{3/2}}\right) + o_p\left(\frac{\rho_{W,N}}{\sqrt{N}}\right). \qquad \textit{Q.E.D.}$$

LEMMA A.18: *Suppose that Assumptions* 1–4, 5(ii), 6, *and* 7 *are satisfied*. *Then the following statements hold*:

(i) $u'P(W)u/N - \tilde{\Lambda}(W)\Sigma_u = O_p(\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N)$.

(ii) $E[h\tilde{\Lambda}(W)\epsilon' v/\sqrt{N}|z] = (K'W/N)\sum_i f_i E(\epsilon_i^2 v_i')/N + O_p(1/N) + O_p(K'W^+/N^2)$.

(iii) $E[hh'\bar{H}^{-1}(W)h/\sqrt{N}|z] = O_p(1/N)$.
(iv) $\sum_{m=1}^{M} w_m E[hh'\bar{H}_m^{-1}f'(I-P_m)\epsilon/N|z] = O_p(1/N)$.
(v) $\sum_{m=1}^{M} w_m E[hh'\bar{H}_m^{-1}v'P_m\epsilon/N|z] = o_p(1/N)$.
(vi) $\sum_{m=1}^{M} w_m E[h\epsilon'(I-P_m)f\bar{H}_m^{-1}f'(I-P_m)\epsilon/N^{-3/2}|z] = O_p(1/N)$.
(vii) $\sum_{m=1}^{M} w_m E[h\epsilon'(I-P_m)f\bar{H}_m^{-1}v'P_m\epsilon/N^{-3/2}|z] = o_p(1/N)$.
(viii) $\sum_{m=1}^{M} w_m E[h\epsilon'P_m v\bar{H}_m^{-1}v'P_m\epsilon/N^{-3/2}|z] = O_p(1/N)$.

PROOF: We begin with the proof of part (i). It holds that $E[\tilde{\Lambda}(W)|z] = \operatorname{tr}(P(W)E[\epsilon\epsilon'])/(N\sigma_\epsilon^2) = (K'W)/N$. We also have

$$
\begin{aligned}
&E\left[\left(\tilde{\Lambda}(W) - \frac{K'W}{N}\right)^2 \middle| z\right] \\
&= \frac{E[\epsilon'P(W)\epsilon\epsilon'P(W)\epsilon|z]}{N^2\sigma_\epsilon^4} - \left(\frac{K'W}{N}\right)^2 \\
&= \frac{\sigma_\epsilon^4(K'W)^2 + 2\sigma_\epsilon^4 W'\Gamma W + O_p\left(\sum_i (P_{ii}(W))^2\right)}{N^2\sigma_\epsilon^4} - \left(\frac{K'W}{N}\right)^2 \\
&= O_p\left(\frac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N^2}\right)
\end{aligned}
$$

by Lemma A.6(iv), replacing u by ϵ, and Lemma A.5(ii). This gives

$$
\left(\tilde{\Lambda}(W) - \frac{K'W}{N}\right)\Sigma_u = O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right).
$$

Similarly, we have

$$
\frac{u'P(W)u}{N} - \frac{K'W}{N}\Sigma_u = O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right).
$$

Thus, part (i) is proved.

For part (ii), we observe that

$$
E\left[\frac{h\tilde{\Lambda}(W)\epsilon'v}{\sqrt{N}}\middle| z\right] = \frac{\sum_{i,j,k,l} E[f_i\epsilon_i\epsilon_j P_{jk}(W)\epsilon_k\epsilon_l v_l'|z]}{N^2\sigma_\epsilon^2}
$$

$$= \frac{\sum_i f_i P_{ii}(W) E[\epsilon_i^4 v_i']}{N^2 \sigma_\epsilon^2} + 2 \frac{\sum_{i \neq j} f_i P_{ij}(W) E[\epsilon_j^2 v_j']}{N^2} + \frac{\sum_{i \neq j} f_i P_{jj}(W) E[\epsilon_i^2 v_i']}{N^2}$$

$$= O_p\left(\frac{K'W^+}{N^2}\right) + o_p\left(\frac{K'W^+}{N^2}\right) + \frac{\sum_{i,j} f_i P_{jj}(W) E[\epsilon_i^2 v_i']}{N^2} - \frac{\sum_i f_i P_{ii}(W) E[\epsilon_i^2 v_i']}{N^2}$$

$$= \frac{K'W}{N} \frac{\sum_i f_i E[\epsilon_i^2 v_i']}{N} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W^+}{N^2}\right),$$

where Lemma A.6(v) implies that

$$\frac{\sum_i f_i P_{ii}(W) E[\epsilon_i^4 v_i']}{N^2 \sigma_\epsilon^2} = O_p\left(\frac{K'W^+}{N^2}\right),$$

$$\frac{\sum_i f_i P_{ii}(W) E[\epsilon_i^2 v_i']}{N^2 \sigma_\epsilon^2} = O_p\left(\frac{K'W^+}{N^2}\right),$$

and the fact that for $f_{a,i}$, the ath element of f_i,

$$\left| \frac{\sum_{i \neq j} f_{a,i} P_{ij}(W)}{N^2} \right| \leq \frac{\sum_{m=1}^{M} |w_m| |(f_a' P_m \mathbf{1}_N)|}{N^2} - \frac{\sum_i f_{a,i} P_{ii}(W)}{N^2}$$

$$\leq \frac{\sum_{m=1}^{M} |w_m| (f_a' P_m f_a)^{1/2} (\mathbf{1}_N' \mathbf{1}_N)^{1/2}}{N^2} + O_p\left(\frac{K'W^+}{N^2}\right)$$

$$\le \left(\frac{f_a'f_a}{N}\right)^{1/2} \frac{\sum_{m=1}^{M} |w_m|}{N} + O_p\left(\frac{K'W^+}{N^2}\right)$$
$$= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W^+}{N^2}\right),$$

gives

$$\frac{\sum_{i\neq j} f_i P_{ij}(W) E[\epsilon_j^2 v_j']}{N^2} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W^+}{N^2}\right).$$

Part (iii) follows Lemma A.8(iii) in Donald and Newey (2001). We have

$$\text{(A.50)} \quad E[hh'\bar{H}^{-1}(W)h/\sqrt{N}|z]$$
$$= \sum_{m=1}^{M} w_m \sum_{i_1,\dots,i_3=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}f_{i_2}'\bar{H}_m^{-1}f_{i_3}\epsilon_{i_3}|z\big]/N^2$$
$$= \sum_{m=1}^{M} w_m \sum_{i=1}^{N} E[\epsilon_i^3|z] f_i f_i' \bar{H}_m^{-1} f_i/N^2 = O_p(1/N).$$

For part (iv), let $\tilde{f}_{i,m}'$ be the ith row of $f'(I-P_m)$ such that

$$E[hh'\bar{H}_m^{-1}f'(I-P_m)\epsilon/N|z] = \sum_{i_1,\dots,i_3=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}f_{i_2}'\bar{H}_m^{-1}\tilde{f}_{i_3,m}\epsilon_{i_3}|z\big]/N^2$$
$$= O_p(1/N)$$

by the same argument as in (A.50).

For part (v), consider

$$E[hh'\bar{H}_m^{-1}v'P_m\epsilon/N|z] = \sum_{i_1,\dots,i_4=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}f_{i_2}'\bar{H}_m^{-1}v_{i_3}P_{m,i_3i_4}\epsilon_{i_4}|z\big]/N^2$$
$$= \sum_{i_1,\dots,i_4=1}^{N} f_{i_1}f_{i_2}'\bar{H}_m^{-1}P_{m,i_3i_4}E\big[\epsilon_{i_1}\epsilon_{i_2}v_{i_3}\epsilon_{i_4}|z\big]/N^2$$
$$= \sum_{i=1}^{N} f_i f_i' \bar{H}_m^{-1} P_{m,ii}\,\mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, \epsilon_i|z]/N^2$$
$$= o_p(N^{-1}).$$

For part (vi), let $\tilde{f}'_{i,m}$ be the ith row of $f'(I-P_m)$ such that

$$E\big[h\epsilon'(I-P_m)f\bar{H}_m^{-1}f'(I-P_m)\epsilon/N^{-3/2}|z\big]$$
$$=\sum_{i_1,\dots,i_3=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}\tilde{f}'_{i_2,m}\bar{H}_m^{-1}\tilde{f}_{i_3,m}\epsilon_{i_3}|z\big]/N^2 = O_p(1/N).$$

For part (vii), consider

$$E\big[h\epsilon'(I-P_m)f\bar{H}_m^{-1}v'P_m\epsilon/N^{-3/2}|z\big]$$
$$=\sum_{i_1,\dots,i_4=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}\tilde{f}'_{i_2,m}\bar{H}_m^{-1}v_{i_3}P_{m,i_3i_4}\epsilon_{i_4}|z\big]/N^2$$
$$=\sum_{i=1}^{N} f_i\tilde{f}'_i\bar{H}_m^{-1}P_{m,ii}\,\mathrm{Cum}[\epsilon_i,\epsilon_i,v_i,\epsilon_i|z]/N^2 = o_p(N^{-1}).$$

For part (viii), consider

$$E\big[h\epsilon'P_m v\bar{H}_m^{-1}v'P_m\epsilon/N^{-3/2}|z\big] \tag{A.51}$$
$$=\sum_{i_1,\dots,i_5=1}^{N} E\big[f_{i_1}\epsilon_{i_1}\epsilon_{i_2}P_{m,i_2i_3}v'_{i_3}\bar{H}_m^{-1}v_{i_4}P_{m,i_4i_5}\epsilon_{i_5}|z\big]/N^2$$
$$=\sum_{i_1,i_2=1}^{N} \sigma_\epsilon^2 f_{i_1}P_{m,i_1i_2}P_{m,i_2i_2}\,\mathrm{tr}\big(\bar{H}_m^{-1}E\big[v_{i_2}v'_{i_2}\epsilon_{i_2}|z\big]\big)/N^2$$
$$+\sum_{i_1,i_2=1}^{N} \sigma_\epsilon^2 f_{i_1}P_{m,i_1i_1}P_{m,i_1i_2}\,\mathrm{tr}\big(\bar{H}_m^{-1}E\big[v_{i_1}v'_{i_1}\epsilon_{i_1}|z\big]\big)/N^2$$
$$+\sum_{i_1,i_2=1}^{N} \sigma_\epsilon^2 f_{i_1}P_{m,i_2i_1}P_{m,i_1i_2}\,\mathrm{tr}\big(\bar{H}_m^{-1}E\big[v_{i_1}v'_{i_1}\epsilon_{i_1}|z\big]\big)/N^2$$
$$+\sum_{i_1,i_2=1}^{N} E\big[\epsilon_{i_2}^3|z\big]f_{i_1}P_{m,i_2i_1}P_{m,i_1i_2}\,\mathrm{tr}\big(\bar{H}_m^{-1}E\big[v_{i_1}v'_{i_1}|z\big]\big)/N^2$$
$$+\sum_{i=1}^{N} f_iP_{m,ii}^2\,\mathrm{tr}(\bar{H}_m^{-1}\,\mathrm{Cum}[\epsilon_i,\epsilon_i,v_i,v'_i,\epsilon_i|z])/N^2,$$

where $E[v_i v_i' \epsilon_i|z]$ does not depend on z by Assumption 3 and, for the first term in (A.51), we have

$$\sum_{i_1,i_2=1}^{N} f_{i_1} P_{m,i_1i_2} P_{m,i_2i_2} = \sum_{i_1,i_2=1}^{N} f_{i_1} P_{m,i_1i_2} = f' P_m \mathbf{1}_N$$
$$\leq (f'f)^{1/2} (\mathbf{1}_N' P_m \mathbf{1}_N)^{1/2} \leq \sqrt{N} (f'f)^{1/2}$$

such that

$$\sum_{i_1,i_2=1}^{N} \sigma_\epsilon^2 f_{i_1} P_{m,i_1i_2} P_{m,i_2i_2} \operatorname{tr}\big(\bar{H}_m^{-1} E\big[v_{i_2} v_{i_2}' \epsilon_{i_2}|z\big]\big)/N^2 = N^{-1}(f'f/N)^{1/2}$$
$$= O_p(N^{-1}),$$

where a similar argument shows that the second term in (A.51) is $O_p(N^{-1})$. Next,

$$\sum_{i_1,i_2=1}^{N} f_{i_1} P_{m,i_2i_1} P_{m,i_1i_2} = \sum_{i_1,i_2=1}^{N} f_{i_1} P_{m,i_1i_2} P_{m,i_2i_1}$$
$$= \sum_{i=1}^{N} f_i P_{m,ii} \leq \sup_i \|f_i\| \sum_{i=1}^{N} P_{m,ii} = O_p(m),$$

where $\sup_i \|f_i\| = O_p(1)$ by Assumption 3(iv) such that the third term in (A.51) is $O_p(m/N^2) = o_p(N^{-1})$ and the same argument also shows that the fourth term in (A.51) is $o_p(N^{-1})$. Finally,

$$\sum_{i=1}^{N} f_i P_{m,ii}^2 \operatorname{tr}(\bar{H}_m^{-1} \operatorname{Cum}[\epsilon_i, \epsilon_i, v_i, v_i', \epsilon_i|z])/N^2$$
$$\leq |\operatorname{tr}(\bar{H}_m^{-1} \operatorname{Cum}[\epsilon_i, \epsilon_i, v_i, v_i', \epsilon_i|z])| \sup_i \|f_i\| \sum_{i=1}^{N} P_{m,ii}^2/N^2$$
$$= o_p(m/N^2) = o_p(N^{-1}).$$

These results establish that $\sum_{m=1}^{M} w_m E[h\epsilon' P_m v \bar{H}_m^{-1} v' P_m \epsilon/N^{-3/2}|z] = O_p(N^{-1})$ as desired. *Q.E.D.*

A.4. *Proof of Theorem A.2*

The MALIML estimator, $\hat{\beta}$ defined in (2.3), has the form

$$\sqrt{N}(\hat{\beta} - \beta_0) = \hat{H}^{-1}\hat{h},$$

$$\hat{H} = X'P(W)X/N - \hat{\Lambda}(W)X'X/N,$$
$$\hat{h} = X'P(W)\epsilon/\sqrt{N} - \hat{\Lambda}(W)X'\epsilon/\sqrt{N}.$$

Also $\hat{H}$ and $\hat{h}$ are decomposed as

$$\hat{h} = h + \sum_{j=1}^{5} T_j^h + Z^h,$$
$$T_1^h = -f'(I - P(W))\epsilon/\sqrt{N}, \quad T_2^h = v'P(W)\epsilon/\sqrt{N},$$
$$T_3^h = -\tilde{\Lambda}(W)\frac{f'\epsilon}{\sqrt{N}}, \quad T_4^h = -\tilde{\Lambda}(W)\frac{v'\epsilon}{\sqrt{N}}, \quad T_5^h = \sqrt{N}\Lambda_q(W)\sigma_{u\epsilon},$$
$$Z^h = (\tilde{\Lambda}(W) - \hat{\Lambda}(W) + \hat{R}_\Lambda)\sqrt{N}\left(\frac{X'\epsilon}{N} - \sigma_{u\epsilon}\right) - \hat{R}_\Lambda\frac{X'\epsilon}{\sqrt{N}}$$

and

$$\hat{H} = H + \sum_{j=1}^{3} T_j^H + Z^H,$$
$$T_1^H = -f'(I - P(W))f/N, \quad T_2^H = (u'f + f'u)/N,$$
$$T_3^H = -\tilde{\Lambda}(W)f'f/N,$$
$$Z^H = u'P(W)u/N - \tilde{\Lambda}(W)\Sigma_u - u'(I - P(W))f/N$$
$$- f'(I - P(W))u/N + \tilde{\Lambda}(W)(H + \Sigma_u) - \hat{\Lambda}(W)X'X/N.$$

Let $T^h = \sum_{j=1}^5 T_j^h$ and $T^H = \sum_{j=1}^3 T_j^H$. We give the order of each term. By Lemma A.5(vi), we have

(A.52) $\quad h = O_p(1) \quad \text{and} \quad H = O_p(1).$

Lemma A.6(ii) gives

(A.53) $\quad T_1^h = O_p\big(\Delta(W)^{1/2}\big).$

By a similar argument to Lemma A.6(iv) (note that $E[v_i\epsilon_i] = 0$), we have

(A.54) $\quad T_2^h = O_p\left(\sqrt{\dfrac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N}}\right).$

Lemma A.17 and the CLT give

$$T_3^h = \left(\frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right)\right) O_p(1) \tag{A.55}$$

$$= O_p\left(\frac{K'W}{N} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right)$$

and

$$T_4^h = O_p\left(\frac{K'W}{N} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right). \tag{A.56}$$

By Lemma A.17, we have

$$T_5^h = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.57}$$

By definition, we have

$$T_1^H = O_p(\Xi(W)), \tag{A.58}$$

where $\Xi(W)$ is defined in (A.11). By a CLT, we have

$$T_2^H = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.59}$$

By Lemmas A.5(vi) and A.17, it holds that

$$T_3^H = \left(\frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right)\right) O_p(1) \tag{A.60}$$

$$= O_p\left(\frac{K'W}{N} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right).$$

By Lemma A.17 together with the CLT, which implies that $\sqrt{N}(\frac{X'\epsilon}{N} - \sigma_{u\epsilon}) = O_p(1)$, as well as

$$\begin{aligned}
&\tilde{\Lambda}(W) - \hat{\Lambda}(W) + \hat{R}_\Lambda \\
&\quad = \left(\frac{\tilde{\sigma}_\epsilon^2}{\sigma_\epsilon^2} - 1\right)\tilde{\Lambda}(W) + \Lambda_q(W) \\
&\quad = O_p\big(N^{-1/2}\big)O_p\left(\frac{K'W}{N} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right) \\
&\qquad + O_p(N^{-1}),
\end{aligned}$$

it follows that

$$\begin{aligned}
Z^h &= O_p\left(\frac{K'W}{N} + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right)\right)O_p\left(\frac{1}{\sqrt{N}}\right) \\
&\quad + O_p\left(\frac{1}{N}\right) + o_p(\rho_{W,N})O_p(1) \\
&= O_p\left(\frac{K'W}{N^{3/2}} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N^{3/2}}\right) \\
&\quad + O_p\left(\frac{1}{N}\right) + o_p(\rho_{W,N}) \\
&= o_p(\rho_{W,N}),
\end{aligned} \tag{A.61}$$

where $1/N = o_p(W'\Gamma W/N) = o_p(\rho_{W,N})$. Last, we have

$$\begin{aligned}
&\tilde{\Lambda}(W)(H + \Sigma_u) - \hat{\Lambda}(W)X'X/N \\
&\quad = \tilde{\Lambda}(W)(H + \Sigma_u - X'X/N) - (\hat{\Lambda}(W) - \tilde{\Lambda}(W))X'X/N \\
&\quad = o_p(\rho_{W,N}) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W}{N^{3/2}}\right) + o_p\left(\frac{\rho_{W,N}}{\sqrt{N}}\right) \\
&\quad = o_p(\rho_{W,N}),
\end{aligned}$$

where $(H + \Sigma_u - X'X/N) = O_p(1/\sqrt{N})$ and $\hat{\Lambda}(W) - \tilde{\Lambda}(W) = O_p(1/N) + O_p(K'W/N^{3/2}) + o_p(\rho_{W,N}/\sqrt{N})$ from Lemma A.17. It then follows that

$$Z^H = o_p(\rho_{W,N}) + O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right) + O_p\left(\frac{\Delta(W)^{1/2}}{\sqrt{N}}\right) \tag{A.62}$$
$$= o_p(\rho_{W,N})$$

by Lemmas A.6(ii), A.17, and A.18(i), the CLT, and the LLN.

We show below that the conditions of Lemma A.1 of Donald and Newey (2001) are satisfied and $S(W)$ has the form given in the theorem.

We first have $h = O_p(1)$ and $H = O_p(1)$ by (A.52). Next, we need to show that $T^h = o_p(1)$. By (A.53)–(A.57), it follows that

$$T^h = O_p\big(\Delta(W)^{1/2}\big) + O_p\left(\sqrt{\frac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N}}\right)$$
$$+ O_p\left(\frac{K'W}{N} + \frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Now Lemma A.6(i) says that $\Delta(W) = o_p(1)$. We have $|K'W/N| \le K'W^+/N \to 0$ by Assumption 5. By Lemma A.6(xii) and Assumption 5, it holds that $W'\Gamma W/N \le CK'W^+/N \to 0$, where C is some constant. Lemma A.5(ii) implies that $\sum_i (P_{ii}(W))^2/N = o_p(K'W^+/N) = o_p(1)$. Thus, $T^h = o_p(1)$ is shown.

The next step is to show that $\|T^H\|^2 = o_p(\rho_{W,N})$. We have, by (A.58)–(A.60), that

$$\|T^H\|^2 = O_p\left(\Xi(W)^2 + \frac{1}{N} + \frac{\Xi(W)}{\sqrt{N}} + \frac{(K'W)^2}{N^2}\right.$$
$$+ \frac{|K'W|}{N}\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}$$
$$+ \frac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N^2} + \Xi(W)\frac{|K'W|}{N}$$

$$+\Xi(W)\frac{\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}}{N}$$
$$+\frac{|K'W|}{N^{3/2}}+\frac{\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}}{N^{3/2}}\Biggr).$$

Since $\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}/N = O_p((W'\Gamma W + \sum_i (P_{ii}(W))^2)/N) = o_p(\rho_{W,N})$, $|K'W|/N^{3/2} = o(|K'W|/N) = o_p(\rho_{W,N})$, $(K'W)^2/N^2 = o(K'W/N) = o_p(\rho_{W,N})$, $1/N = o_p(\rho_{W,N})$, and the observation that $\Xi(W)/\sqrt{N} = o_p(\rho_{W,N})$ by Lemma A.6(vi) and $\Xi(W) = O_p(\Delta(W)^{1/2})$, we have

$$\|T^H\|^2 = O_p\bigl((\Xi(W))^2\bigr) + o_p(\rho_{W,N}).$$

The order of $(\Xi(W))^2$ is $o_p(\rho_{W,N})$ by Lemma A.7. Next, we consider $\|T^h\|\cdot\|T^H\|$. We have, by (A.53)–(A.60),

$$\begin{aligned}
&\|T^h\|\cdot\|T^H\| \\
&= O_p\Biggl(\Delta(W)^{1/2}+\sqrt{\frac{W'\Gamma W+\sum_i (P_{ii}(W))^2}{N}}+\frac{|K'W|}{N} \\
&\quad +\frac{\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}}{N}+\frac{1}{\sqrt{N}}\Biggr) \\
&\quad \cdot O_p\Biggl(\Xi(W)+\frac{1}{\sqrt{N}}+\frac{|K'W|}{N}+\frac{\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}}{N}\Biggr) \\
&= O_p\Biggl(\Delta(W)^{1/2}\Xi(W)+\frac{\Delta(W)^{1/2}}{\sqrt{N}}+\frac{\sqrt{W'\Gamma W+\sum_i (P_{ii}(W))^2}}{N} \\
&\quad +\Xi(W)\sqrt{\frac{W'\Gamma W+\sum_i (P_{ii}(W))^2}{N}}\Biggr) \\
&\quad + o_p(\rho_{W,N})
\end{aligned}$$

$$
\begin{aligned}
&= O_p\left(\Delta(W)^{1/2}\Xi(W) + \Xi(W)\sqrt{\frac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N}}\right) \\
&\quad + o_p(\rho_{W,N}) \\
&= o_p(\rho_{W,N}),
\end{aligned}
$$

since $o_p(1)|K'W|/N = o_p(\rho_{W,N})$, $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N = o_p(\rho_{W,N})$, $1/N = o_p(\rho_{W,N})$, $\Delta(W)^{1/2}/\sqrt{N} = o_p(\rho_{W,N})$ by Lemma A.6(vi), and the order of $\Xi(W)$ is $o_p(\Delta(W)^{1/2})$ by Lemma A.7. Last, it holds that $Z^h = o_p(\rho_{W,N})$ and $Z^H = o_p(\rho_{W,N})$ by (A.61) and (A.62).

We have shown that the conditions of Lemma A.1 of Donald and Newey (2001)[2] are satisfied and we apply the lemma with

$$
\begin{aligned}
\hat{A}(W) &= (h + T_1^h + T_2^h)(h + T_1^h + T_2^h)' + h(T_3^h + T_4^h + T_5^h)' \\
&\quad + (T_3^h + T_4^h + T_5^h)h' - hh'H^{-1}(T_1^H + T_2^H + T_3^H) \\
&\quad - (T_1^H + T_2^H + T_3^H)H^{-1}hh'
\end{aligned}
$$

and

$$
\begin{aligned}
Z^A(W) &= (T_3^h + T_4^h + T_5^h)(T_3^h + T_4^h + T_5^h)' \\
&\quad + (T_3^h + T_4^h + T_5^h)(T_1^h + T_2^h)' + (T_1^h + T_2^h)(T_3^h + T_4^h + T_5^h).
\end{aligned}
$$

We show that $Z^A(W) = o_p(\rho_{W,N})$. By (A.55), (A.56), and the fact that $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N = o_p(\rho_{W,N})$, it holds that

$$
\begin{aligned}
(T_3^h + T_4^h)(T_3^h + T_4^h)' &= O_p\left(\left(\frac{K'W}{N}\right)^2\right) + o_p(\rho_{W,N}) \\
&= o_p(\rho_{W,N}).
\end{aligned}
$$

By (A.55), (A.56), (A.57), and the fact that $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N^{3/2} = o_p(\rho_{W,N})$, it holds that

$$
T_5^h(T_3^h + T_4^h)' = O_p\left(\frac{K'W}{N^{3/2}}\right) + o_p(\rho_{W,N}) = o_p(\rho_{W,N}).
$$

[2]We note that here we do not need to use our Lemma A.1, which is a modified version of Lemma A.1 of Donald and Newey (2001).

By (A.57), we have

$$T_5^h(T_5^h)' = O_p\left(\frac{1}{N}\right) = o_p(\rho_{W,N}).$$

By (A.53), (A.55), (A.56), and the fact that $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N = o_p(\rho_{W,N})$, we have

$$T_1^h(T_3^h + T_4^h) = O_p\left(\Delta(W)^{1/2}\frac{K'W}{N}\right) + o_p(\rho_{W,N}) = o_p(\rho_{W,N}),$$

since $\Delta(W)^{1/2} = o_p(1)$ by Lemma A.6(i). By (A.54), (A.55), (A.56), and the fact that $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N = o_p(\rho_{W,N})$, it follows that

$$\begin{aligned} T_2^h(T_3^h + T_4^h) &= O_p\left(\frac{K'W}{N}\sqrt{\frac{W'\Gamma W + \sum_i (P_{ii}(W))^2}{N}}\right) + o_p(\rho_{W,N}) \\ &= o_p(\rho_{W,N}). \end{aligned}$$

Lemma A.6(vi), (A.53), and (A.57) imply that

$$T_5^h(T_1^h)' = O_p\left(\frac{\Delta(W)^{1/2}}{\sqrt{N}}\right) = o_p(\rho_{W,N}).$$

Last, we have

$$T_5^h(T_2^h)' = O_p\left(\frac{\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}}{N}\right) = o_p(\rho_{W,N})$$

by (A.54), (A.57), and the fact that $\sqrt{W'\Gamma W + \sum_i (P_{ii}(W))^2}/N = o_p(\rho_{W,N})$. To sum up, we have $Z^A(W) = o_p(\rho_{W,N})$.

Now, we calculate the expectation of each term in $\hat{A}(W)$. First of all, $E[hh'|z] = E[f'\epsilon\epsilon' f/N|z] = \sigma_\epsilon^2 H$. Second, we have

$$E[hT_1^{h\prime}|z] = E\left[-\frac{f'\epsilon\epsilon'(I - P(W))f}{N}\Big|z\right] = -\sigma_\epsilon^2\frac{f'(I - P(W))f}{N}.$$

Similarly, it holds that $E[T_1^h h'|z] = -\sigma_\epsilon^2 f'(I - P(W))f/N$. Third, using Lemma A.6(v) and replacing u by v gives

$$E[hT_2^{h\prime}|z] = \sum_{i=1}^N f_i P_{ii}(W)E[\epsilon_i^2 v_i'|z]/N,$$

which is $O_p(K'W^+/N)$. Fourth,

$$E[T_1^h T_1^{h\prime}|z] = E\left[\frac{f'(I-P(W))\epsilon\epsilon'(I-P(W))f}{N}\Big|z\right]$$
$$= \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N}.$$

Fifth, by Lemma A.6(viii), replacing u by v, we obtain

$$E[T_1^h T_2^{h\prime}|z] = -E\left[\frac{f'(I-P(W))\epsilon\epsilon' P(W)v}{N}\Big|z\right]$$
$$= -\frac{f'(I-P(W))\mu_v(W)}{N},$$

where $\mu_v(W) = (\mu_{v,1}(W), \ldots, \mu_{v,N}(W))$ and $\mu_{v,i} = P_{ii}(W)E[\epsilon_i^2 v_i]$. Similarly, we have $E[T_2^h T_1^{h\prime}|z] = -\mu_v(W)(I-P(W))f/N$. Sixth, noting that $E[v_i\epsilon_i|z] = 0$, a similar argument to Lemma A.6(iv) gives

$$E[T_2^h T_2^{h\prime}|z] = \sigma_\epsilon^2 \Sigma_v (W'\Gamma W)/N + \mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, v_i'] \sum_i (P_{ii}(W))^2/N.$$

Seventh, we have

$$E[hh'H^{-1}T_1^{H\prime}|z] = -E\left[\frac{f'\epsilon\epsilon' f H^{-1} f'(I-P(W))f}{N^2}\Big|z\right]$$
$$= -\sigma_\epsilon^2 \frac{f'(I-P(W))f}{N}.$$

Similarly, we have $E[T_1^H H^{-1}hh'|z] = -\sigma_\epsilon^2 f'(I-P(W))f/N$. Eighth, Lemma A.6(vii) implies that

$$E[hh'H^{-1}T_2^{H\prime}|z] = E\left[\frac{hh'H^{-1}(u'f+f'u)}{N}\Big|z\right]$$
$$= O_p\left(\frac{1}{N}\right) = o_p(\rho_{W,N})$$

and that $E[T_2^H H^{-1}hh'|z] = o_p(\rho_{W,N})$. Ninth, we have

$$h(T_3^h)' - hh'H^{-1}(T_3^H)' = T_3^h h' - T_3^H H^{-1}hh' = 0.$$

Tenth, we have

$$E[h(T_4^h)'|z] = -\frac{K'W}{N}\frac{\sum_{i=1}^{N} f_i E[\epsilon_i^2 v_i']}{N} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{K'W^+}{N^2}\right)$$

$$= -\frac{K'W}{N}\frac{\sum_{i=1}^{N} f_i E[\epsilon_i^2 v_i']}{N} + o_p(\rho_{W,N})$$

by Lemma A.18(ii). Similarly, we have $E[T_4^h h'|z] = -(K'W/N)(\sum_{i=1}^N f_i \times E[\epsilon_i^2 u_i]/N) + o_p(\rho_{W,N})$. Last, Lemma A.18(iii)–(viii) implies that

$$E[h(T_5^h)'|z] = O_p\left(\frac{1}{N}\right) = o_p(\rho_{W,N})$$

and that $E[T_5^h h'|z] = o_p(\rho_{W,N})$.

Let

$$\hat{\zeta} = \sum_{i=1}^{N} f_i P_{ii}(W) E[\epsilon_i^2 v_i']/N - \frac{K'W}{N}\sum_{i=1}^{N} f_i E[\epsilon_i^2 v_i']/N.$$

Note that $\hat{\zeta} = 0$ under the third moment condition in the theorem. Therefore, we have

$$\begin{aligned}
E(\hat{A}(K)) &= \sigma_\epsilon^2 H - \sigma_\epsilon^2 \frac{f'(I-P(W))f}{N} - \sigma_\epsilon^2 \frac{f'(I-P(W))f}{N} \\
&\quad - \frac{f'(I-P(W))\mu_v(W)}{N} - \frac{\mu_v(W)'(I-P(W))f}{N} \\
&\quad + \sigma_\epsilon^2 \Sigma_v \frac{W'\Gamma W}{N} + \mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, v_i'] \frac{\sum_i (P_{ii}(W))^2}{N} \\
&\quad + \hat{\zeta} + \hat{\zeta}' + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N} \\
&\quad + \sigma_\epsilon^2 \frac{f'(I-P(W))f}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))f}{N} + o_p(\rho_{W,N}) \\
&= \sigma_\epsilon^2 H + \sigma_\epsilon^2 \Sigma_v \frac{W'\Gamma W}{N} + \sigma_\epsilon^2 \frac{f'(I-P(W))(I-P(W))f}{N} \\
&\quad + \mathrm{Cum}[\epsilon_i, \epsilon_i, v_i, v_i'] \frac{\sum_i (P_{ii}(W))^2}{N}
\end{aligned}$$

$$
\begin{aligned}
&-\frac{f'(I-P(W))\mu_v(W)}{N}-\frac{\mu_v(W)'(I-P(W))f}{N}\\
&+\hat{\zeta}+\hat{\zeta}'+o_p(\rho_{W,N}).
\end{aligned}
$$

By Lemma A.1 of Donald and Newey (2001), we have the desired result.

For the MAFuller estimator $\hat{\beta}$ defined in (2.4) the result can be established by noting the following. By the construction of $\hat{\Lambda}_m$, we have $0 \leq 1-\hat{\Lambda}_m \leq 1$. Therefore,

$$
\begin{aligned}
0 &< \check{\Lambda}_m - \hat{\Lambda}_m\\
&= \frac{\dfrac{\alpha}{N-m}(1-\hat{\Lambda}_m)^2}{1-\dfrac{\alpha}{N-m}(1-\hat{\Lambda}_m)}\\
&= \frac{\alpha((1-\hat{\Lambda}_m)^2)}{N-m-\alpha(1-\hat{\Lambda}_m)}\\
&\leq \frac{\alpha}{N-M-\alpha} = O\left(\frac{1}{N}\right)
\end{aligned}
$$

uniformly over m. It therefore follows that

$$
\check{\Lambda}(W) = \hat{\Lambda}(W) + O_p(1/N). \tag{A.63}
$$

Now let $\rho_{W,N} = \operatorname{tr}(S(W))$. We have

$$
\begin{aligned}
\frac{X'P(W)X}{N} - \check{\Lambda}(W)\frac{X'X}{N} &= \frac{X'P(W)X}{N} - \hat{\Lambda}(W)\frac{X'X}{N} + O_p\left(\frac{1}{N}\right)\\
&= \frac{X'P(W)X}{N} - \hat{\Lambda}(W)\frac{X'X}{N} + o_p(\rho_{W,N})
\end{aligned}
$$

by (A.63), $X'X/N = O_p(1)$ and $1/N = o_p(\rho_{W,N})$. Similarly, we have

$$
\begin{aligned}
\frac{X'P(W)\epsilon}{\sqrt{N}} - \check{\Lambda}(W)\frac{X'\epsilon}{\sqrt{N}} &= \frac{X'P(W)\epsilon}{\sqrt{N}} - \hat{\Lambda}(W)\frac{X'\epsilon}{\sqrt{N}} + O_p\left(\frac{1}{N}\right)\\
&= \frac{X'P(W)\epsilon}{\sqrt{N}} - \hat{\Lambda}(W)\frac{X'\epsilon}{\sqrt{N}} + o_p(\rho_{W,N}).
\end{aligned}
$$

Therefore, the higher order mean squared errors of the MALIML and the MAFuller estimator are the same. *Q.E.D.*

A.5. *Verification of Regularity Conditions for Unconstrained Optimal Weights*

To demonstrate that the regularity conditions imposed are not too stringent, it is useful to consider various optimal weights and verify that the conditions hold. We note that when Ω is equal to Ω_U or Ω_B, a closed form solution for W^* is available. Let $\gamma_m = \lambda' H^{-1} f'(I - P_m) f H^{-1}\lambda/N$ and let U be the matrix whose (i, j) element is $\gamma_{\max(i,j)}$ so that $\lambda' H^{-1} f'(I - P(W))(I - P(W)) f H^{-1}\lambda/N = W'UW$. This implies that $S_\lambda(W)$ is quadratic function in W and the optimal weight is given by solving the first-order condition. For the MALIML estimator with $\Omega = \Omega_U$, we have

$$W^* = (\mathbf{1}_M'(U + \sigma_v^2\Gamma)^{-1}\mathbf{1}_M)^{-1}(U + \sigma_v^2\Gamma)^{-1}\mathbf{1}_M$$

$$= \begin{pmatrix} \dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_1 - \gamma_2)} \\ -\dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_1 - \gamma_2)} + \dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_2 - \gamma_3)} \\ \vdots \\ -\dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_{M-2} - \gamma_{M-1})} + \dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_{M-1} - \gamma_M)} \\ -\dfrac{\sigma_v^2}{\sigma_v^2 + N(\gamma_{M-1} - \gamma_M)} + 1 \end{pmatrix}$$

such that

$$\sum_{s=1}^{j} w_s = \frac{\sigma_v^2}{\sigma_v^2 + N(\gamma_j - \gamma_{j+1})}.$$

It follows that for some $\varepsilon > 0$,

$$\left|\sum_{s=1}^{j} w_s\right| \le \frac{j^{2\alpha+1}}{N}\frac{\sigma_v^2}{j^{2\alpha+1}\sigma_v^2/N + \varepsilon} \quad \text{wpa1 for } j \notin \bar{J}$$

and

$$\left|\sum_{s=1}^{j} w_s\right| \le \frac{L^{2\alpha+1}}{N}\frac{\sigma_v^2}{\varepsilon} \quad \text{for} \quad j \notin \bar{J}, j \le L$$

such that, for $L = O(N^{1/(2(2\alpha+1))})$, it follows that

$$\sup_{j\notin\bar{J}, j\le L}\left|\sum_{s=1}^{j} w_s\right| = O_p(1/\sqrt{N}).$$

The case of MA2SLS with $\Omega = \Omega_U$ is handled next. The optimal weight is given by

$$
\begin{aligned}
W_U^* &= \arg\min_{W \in \Omega_U} S_\lambda(W) \\
&= \frac{1}{2} A^{-1}\left(K\lambda' H^{-1} B_N H^{-1}\lambda + \frac{2 - \mathbf{1}_M' A^{-1} K \lambda' H^{-1} B_N H^{-1}\lambda}{\mathbf{1}_M' A^{-1}\mathbf{1}_M}\mathbf{1}_M \right) \\
&= e_M + \frac{1}{2}\frac{2(\sigma_\epsilon^2\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2 + M\sigma_{\lambda\epsilon}^2) - B_\lambda}{\sigma_\lambda^2\sigma_\epsilon^2 + \sigma_{\lambda\epsilon}^2 + \sigma_{\lambda\epsilon}^2 \sum_{j=1}^{M-1} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_j - \gamma_{j+1})}} \\
&\quad \times \begin{pmatrix} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_1 - \gamma_2)} \\ -\frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_1 - \gamma_2)} + \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_2 - \gamma_3)} \\ \vdots \\ -\frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_{M-1} - \gamma_M)} \end{pmatrix}.
\end{aligned}
$$

First, consider $(\gamma_j - \gamma_{j+1}) < \gamma_j \leq C < \infty$ which holds because f_i is bounded by Assumption 3(iv) such that $\lambda' H^{-1} f'(I - P_m) f H^{-1}\lambda/N \leq \lambda' H^{-1} f' \times f H^{-1}\lambda/N \leq \sup_i \|f_i\|^2 \|\lambda' H^{-1}\|^2 \leq C$ for some $C < \infty$. Then

$$
\begin{aligned}
&\sum_{j=1}^{M-1} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_j - \gamma_{j+1})} \\
&\quad \geq \frac{1}{N}\sum_{j=1}^{M-1} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{(\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2)/N + C} \\
&\quad = O\left(\frac{M}{N}\right)
\end{aligned}
$$

such that

$$
\frac{2(\sigma_\epsilon^2\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2 + M\sigma_{\lambda\epsilon}^2) - B_\lambda}{\sigma_\lambda^2\sigma_\epsilon^2 + \sigma_{\lambda\epsilon}^2 + \sigma_{\lambda\epsilon}^2 \sum_{j=1}^{M-1} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_j - \gamma_{j+1})}} = O_p(M).
$$

By the same argument as before we have

$$\frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + N(\gamma_j - \gamma_{j+1})} \le \frac{L^{2\alpha+1}}{N} \frac{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2}{\sigma_\lambda^2 + \sigma_{\lambda\epsilon}^2/\sigma_\epsilon^2 + \varepsilon} \quad \text{for} \quad j \notin \bar{J}, j \le L$$

such that

$$\sup_{j\notin\bar{J}, j\le L} \left| \sum_{s=1}^{j} w_s \right| = O_p\left(\frac{ML^{2\alpha+1}}{N} \right),$$

where the desired rate obtains if

$$L = o\left(\left(\frac{N^{1/2}}{M} \right)^{1/(2\alpha+1)} \right).$$

REFERENCES

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

FULLER, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–954.

HANSEN, B. E. (2007): "Least Squares Model Averaging," *Econometrica*, 75, 1175–1189.

LEWIS, R., AND G. C. REINSEL (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

LI, K.-C. (1987): "Asymptotic Optimality for C_p, C_L, Cross-Validation and Generalized Cross-Validation: Discrete Index Set," *The Annals of Statistics*, 15, 958–975.

MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus With Applications in Statistics and Econometrics*. Chichester: Wiley.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.

WHITTLE, P. (1960): "Bounds for the Moments of Linear and Quadratic Forms in Independent Variables," *Theory of Probability and its Applications*, 5, 302–305.

Dept. of Economics, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, U.S.A.; gkuerste@ucdavis.edu

and

Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan; okui@kier.kyoto-u.ac.jp.

Manuscript received September, 2007; final revision received October, 2009.