Supplementary Material

for "Unbalanced Spatial Panel Data Models with Fixed Effects"

Xiaoyu Meng and Zhenlin Yang

In this Supplementary Material, we provide detailed proofs of the results not given in the main text due to space constraint. We also provide some details on the immediate extensions of the proposed AQS method.

Additional Proofs

Proof of Lemma A.3: Proof is simpler using a $\mathbf{D}_\alpha^\star$ under the constraint $\alpha_1 = 0$.

Proof of (i)**.** Let $\mathbb{D}_\mu(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\mu$, $\mathbb{D}_\alpha(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\alpha^\star$, $\mathcal{D}_{11}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{D}_\mu(\rho)$, $\mathcal{D}_{12}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{D}_\alpha(\rho)$, $\mathcal{D}_{22}(\rho) = \mathbb{D}_\alpha'(\rho)\mathbb{D}_\alpha(\rho)$ and $\mathcal{F}(\rho) = \mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)$. Using the inverse formula of a partioned matrix, one has

$$
\begin{aligned}
[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1} &= \begin{bmatrix} \mathcal{D}_{11}(\rho) & \mathcal{D}_{12}(\rho) \\ \mathcal{D}_{12}'(\rho) & \mathcal{D}_{22}(\rho) \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \mathcal{F}^{-1}(\rho) & -\mathcal{F}^{-1}(\rho)\mathcal{D}_{12}(\rho)\mathcal{D}_{22}^{-1}(\rho) \\ -\mathcal{D}_{22}^{-1}(\rho)\mathcal{D}_{12}'(\rho)\mathcal{F}^{-1}(\rho) & \mathcal{D}_{22}^{-1}(\rho) + \mathcal{D}_{22}^{-1}(\rho)\mathcal{D}_{12}'(\rho)\mathcal{F}^{-1}(\rho)\mathcal{D}_{12}(\rho)\mathcal{D}_{22}^{-1}(\rho) \end{bmatrix}.
\end{aligned}
$$

Plugging this into $\mathbb{Q}_\mathbb{D}(\rho)$, we obtain after some algebra,

$$
\mathbb{Q}_\mathbb{D}(\rho) = \mathbb{Q}_{\mathbb{D}_\alpha}(\rho) - \mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)[\mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)\mathbb{D}_\mu(\rho)]^{-1}\mathbb{D}_\mu'(\rho)\mathbb{Q}_{\mathbb{D}_\alpha}(\rho). \tag{D.1}
$$

Given the special structure of $\mathbb{D}_\alpha(\rho)$, one has $\mathbb{Q}_{\mathbb{D}_\alpha}(\rho) = \texttt{blkdiag}(J_1(\rho), \ldots, J_T(\rho))$, where $J_1(\rho) = I_{n_1}$ and $J_t(\rho) = I_{n_t} - \frac{1}{n_t}B_t(\rho)l_{n_t}[\frac{1}{n_t}l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}]^{-1}l_{n_t}'B_t'(\rho)$ for $t = 2, \cdots, T$. By Assumption D, the limit of $\frac{1}{n_t}l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}$ is bounded away from zero and the elements of $B_t(\rho)l_{n_t}l_{n_t}'B_t'(\rho)$ are uniformly bounded, uniformly in $\rho \in \Delta_\rho$ for each t. Therefore, $J_t(\rho)$ must be uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$ for all t. Hence, $\mathbb{Q}_{\mathbb{D}_\alpha}(\rho)$ is also uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$.

We next consider the second term on the RHS of equation (D.1). We denote it as $\bar{\mathcal{Q}}(\rho)$, which can be partitioned into $T \times T$ blocks with (s,t)th block being

$$
\bar{\mathcal{Q}}_{s,t}(\rho) = -\tfrac{1}{T}J_s(\rho)B_s(\rho)D_s[\tfrac{1}{T}\textstyle\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)J_t(\rho).
$$

By assuming $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)$ is uniformly bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$, for all s and t, we have that the row and column sums of each $\bar{\mathcal{Q}}_{s,t}(\rho)$ must have uniform order $O(1/T)$, uniformly in $\rho \in \Delta_\rho$. As there are T blocks in each row or in each column of $\bar{\mathcal{Q}}(\rho)$, we must have $\bar{\mathcal{Q}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. Consequently, $\mathbb{Q}_{\mathbb{D}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$.

Proof of (ii)**.** Let $Z_N(\rho) = [\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ with its (j,k)th element being denoted by $z_{jk}(\rho)$. From Assumption C(ii), $Z_N(\rho)$ converges to a finite limit uniformly in $\rho \in \Delta_\rho$. Therefore, there exists a constant c_z such that $|z_{jk}(\rho)| \leq c_z$ uniformly in $\rho \in \Delta_\rho$ for large enough N. Note that $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$. As the elements of $\mathbf{X}$ are uniformly bounded (Assumption C(i)), and $\mathbf{B}_N(\rho)$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ are bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$, the elements of $\mathbb{X}(\rho)$ are also uniformly bounded, uniformly in $\rho \in \Delta_\rho$. Hence, there exists a constant c_x such that $|x_{jk}(\rho)| \leq c_x$ uniformly in $\rho \in \Delta_\rho$, where $x_{jk}(\rho)$ is the (j,k)th element of $\mathbb{X}(\rho)$. Let $p_{jl}(\rho)$ be the (j,l)th element of $\mathbb{P}_{\mathbb{X}}(\rho) = \frac{1}{N}\mathbb{X}(\rho)[\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)$. It follows that uniformly in $\rho \in \Delta_\rho$, $\sum_{j=1}^N |p_{jl}(\rho)| \leq \frac{1}{N}\sum_{j=1}^N \sum_{r=1}^k \sum_{s=1}^k |z_{rs}(\rho)x_{jr}(\rho)x_{ls}(\rho)| \leq k^2 c_z c_x^2$ for all $l = 1, 2, \ldots, N$. Similarly, uniformly in $\rho \in \Delta_\rho$, we have $\sum_{l=1}^N |p_{jl}(\rho)| \leq \frac{1}{N}\sum_{l=1}^N \sum_{r=1}^k \sum_{s=1}^k |z_{rs}(\rho)x_{jr}(\rho)x_{ls}(\rho)| \leq k^2 c_z c_x^2$ for all $j = 1, 2, \ldots, N$. That is, $\mathbb{P}_{\mathbb{X}}(\rho)$ is bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. Consequently, $\mathbb{Q}_{\mathbb{X}}(\rho) = I_N - \mathbb{P}_{\mathbb{X}}(\rho)$ is also bounded in both row and column sum norms, uniformly in $\rho \in \Delta_\rho$. ∎

Proof of Lemma A.4: From the proof of Lemma A.3, the elements of $\mathbb{X}(\rho)$ and the elements of $[\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ are uniformly bounded, uniformly in $\rho \in \Delta_\rho$. If A_N and B_N are bounded in row (column) sum norm, then $A_N B_N$ is also bounded in row (column) sum norm. Thus, Lemma A.6 of Lee (2004b) implies that the elements of $\frac{1}{N}\mathbb{X}'(\rho)A_N B_N \mathbb{X}(\rho)$ are uniformly bounded. It follows that $\mathtt{tr}[A_N \mathbb{P}_{\mathbb{X}}(\rho) B_N] = \mathtt{tr}[(\frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho))^{-1}\frac{1}{N}\mathbb{X}'(\rho)A_N B_N \mathbb{X}(\rho)] = O(1)$, uniformly in $\rho \in \Delta_\rho$ because the number of regressors k is fixed. ∎

Proof of Lemma A.6: Firstly, Lemma A.8 of Lee (2004b) implies that $\mathtt{tr}(\mathbf{H}A_N)$, $\mathtt{tr}(A_N A_N')$, $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}A_N)$ and $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}A_N')$ are all $O(\frac{N}{h_n})$. As $\sum_{i=1}^N a_{ii}^2 \leq \mathtt{tr}(A_N A_N')$, we also have $\sum_{i=1}^N a_{ii}^2 = O(\frac{N}{h_n})$. These and Lemma A.5 show that $\mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = \mathtt{tr}(\mathbf{H}A_N) = O(\frac{N}{h_n})$ and $\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = \sum_{i=1}^N a_{ii}^2[\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}[\mathbf{H}A_N(\mathbf{H}A_N' + \mathbf{H}A_N)] = O(\frac{N}{h_n})$. As $\mathrm{E}[(\mathbf{V}'A_N\mathbf{V})^2] = \mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) + \mathrm{E}^2(\mathbf{V}'A_N\mathbf{V}) = O((\frac{N}{h_n})^2)$, we have $P(\frac{h_n}{N}|\mathbf{V}'A_N\mathbf{V}| \geq M) \leq \frac{1}{M^2}(\frac{h_n}{N})^2\mathrm{E}[(\mathbf{V}'A_N\mathbf{V})^2] = O(1)$, by the generalized Chebyshev's inequality. It follows that $\mathbf{V}'A_N\mathbf{V} = O_p(\frac{N}{h_n})$. Moreover, by Chebyshev's inequality, $P((\frac{h_n}{N})^{\frac{1}{2}}|\mathbf{V}'A_N\mathbf{V} - \mathrm{E}(\mathbf{V}'A_N\mathbf{V})| \geq$

$M) \leq \frac{1}{M^2}\frac{h_n}{N}\mathrm{Var}(\mathbf{V}'A_N\mathbf{V}) = O(1)$. This implies that $\mathbf{V}'A_N\mathbf{V} - \mathrm{E}(\mathbf{V}'A_N\mathbf{V}) = O_p((\frac{N}{h_n})^{\frac{1}{2}})$. Finally, as the elements of c_N have uniform order $O(h_n^{-1/2})$, there exists a constant $\bar{c}$ such that $|c_j| \leq \frac{\bar{c}}{h_n^{1/2}}$ for all j. Hence, we have by the boundedness of $\|A_N\|_1$,

$$\begin{aligned}\mathrm{Var}[(\tfrac{h_n}{N})^{\frac{1}{2}}c_N'A_N\mathbf{V}] &= \tfrac{h_n}{N}\textstyle\sum_{i=1}^N\sum_{j=1}^N\sum_{k=1}^N c_jc_ka_{ji}a_{ki}\sigma_i^2\\ &\leq \bar{c}^2(\tfrac{1}{N}\textstyle\sum_{i=1}^N\sigma_i^2)(\sum_{j=1}^N|a_{ji}|)(\sum_{k=1}^N|a_{ki}|) = O(1).\end{aligned}$$

It follows that $c_N'A_N\mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$, by Chebyshev's inequality. ■

Proof of Corollary 2.1: Note that $\Gamma_N^*(\hat{\theta}_N^*) = \Gamma_N^*(\theta)|_{(\theta=\hat{\theta}_N^*,\phi=\hat{\phi}_N^*,\gamma=\hat{\gamma}_N,\kappa=\hat{\kappa}_N)}$. As $\hat{\theta}_N^*$, $\hat{\gamma}_N$ and $\hat{\kappa}_N$ are consistent estimators for θ_0, γ and κ, plugging these estimators into $\Gamma_N^*(\theta)$ will not bring additional bias to the estimation of $\Gamma_N^*(\theta_0)$. However, due to incidental parameters problem, the $\hat{\mu}_N^*$ component of $\hat{\phi}_N^*$ is not consistent for the estimation of μ_0 when T is fixed. The estimation bias caused by replacing ϕ_N by $\hat{\phi}_N^*$ can be derived as follow. Recall (2.4),

$$\hat{\phi}_N(\beta,\delta) = [\mathbb{D}_N'(\rho)\mathbb{D}_N(\rho)]^{-1}\mathbb{D}_N'(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta].$$

Thus, the unconstrained estimate of ϕ_0 is just $\hat{\phi}_N^* = \hat{\phi}_N(\hat{\beta}_N^*,\hat{\delta}_N^*)$. Note $\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^* = \mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0 - \mathbf{W}\mathbf{Y}(\hat{\lambda}_N^* - \lambda_0) - \mathbf{X}(\hat{\beta}_N^* - \beta_0)$. Applying the MVT on each row of $\mathbf{D}\hat{\phi}_N^*$ with respect to the $\hat{\rho}_N^*$-element, we have,

$$\begin{aligned}\mathbf{D}\hat{\phi}_N^* &= \mathbf{D}[\mathbb{D}_N'(\hat{\rho}_N^*)\mathbb{D}_N(\hat{\rho}_N^*)]^{-1}\mathbb{D}_N'(\hat{\rho}_N^*)\mathbf{B}_N(\hat{\rho}_N^*)[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*] \qquad \text{(D.2)}\\ &= \mathbf{B}_N^{-1}(\hat{\rho}_N^*)\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^*)\mathbf{B}_N(\hat{\rho}_N^*)[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*]\\ &= [\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_N - \mathbb{R}_N(\bar{\rho})(\hat{\rho}_N^* - \rho_0)][\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*]\\ &= \mathbf{D}\phi_0 + \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}_N^* - \lambda_0) + \mathbf{X}(\hat{\beta}_N^* - \beta_0)]\\ &\quad - \mathbb{R}_N(\bar{\rho})[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*](\hat{\rho}_N^* - \rho_0),\end{aligned}$$

where $\bar{\rho}$ lies between $\hat{\rho}_N^*$ and ρ_0 and changes over the rows of $\mathbb{R}_N(\bar{\rho})$, and $\mathbb{R}_N(\rho)$ is given below (B.4). From its expression, $\Gamma_N^*(\theta)$ is seen to have components that are either linear or quadratic in $\mathbf{D}\phi$. Let d_N be a non-stochastic N-vector with elements being of uniform order $O(1)$ or $O(h_n^{-1})$. Using (D.2), the terms of $\Gamma_N^*(\hat{\theta}_N^*)$ linear in $\mathbf{D}\hat{\phi}_N^*$ can be represented as

$$\begin{aligned}\tfrac{1}{N_1}d_N'\mathbf{D}\hat{\phi}_N^* &= \tfrac{1}{N_1}d_N'\mathbf{D}\phi_0 + \tfrac{1}{N_1}d_N'\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \tfrac{1}{N_1}d_N'\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}_N^* - \lambda_0) + \mathbf{X}(\hat{\beta}_N^* - \beta_0)]\\ &\quad + \tfrac{1}{N_1}d_N'\mathbb{R}_N(\bar{\rho})[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*](\hat{\rho}_N^* - \rho_0) = \tfrac{1}{N_1}d_N'\mathbf{D}\phi_0 + o_p(1),\end{aligned}$$

where the last equation holds because of the consistency of $\hat{\theta}_N^*$ and Lemma A.6, using $\mathbf{Y} =$

$\mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$. Hence, we can conclude that the terms of $\Gamma_N^*(\theta_0)$ linear in ϕ_0 can be consistently estimated by simply replacing ϕ_0 with $\hat{\phi}_N^*$.

The only term that is quadratic in ϕ_0 is contained in $\Gamma_{\lambda\lambda}^*(\theta_0)$, which is $\frac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbb{D}_N'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}_N\phi_0$. The plug-in estimator estimates this term by $\frac{1}{N_1\hat{\sigma}_{v,N}^{*2}}\hat{\phi}_N^{*\prime}\mathbb{D}_N'(\hat{\rho}_N^*)\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbb{D}_N(\hat{\rho}_N^*)\hat{\phi}_N^*$. Using (D.2), $\hat{\theta}_N^* - \theta_0 = o_p(1)$ and Lemma A.6, we show that this estimator is biased/inconsistent:

$$
\begin{aligned}
&\tfrac{1}{N_1\hat{\sigma}_{v,N}^{*2}}\hat{\phi}_N^{*\prime}\mathbb{D}_N'(\hat{\rho}_N^*)\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbb{D}_N(\hat{\rho}_N^*)\hat{\phi}_N^*\\
&= \tfrac{1}{N_1\hat{\sigma}_{v,N}^{*2}}\phi_0'\mathbb{D}_N'(\hat{\rho}_N^*)\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbb{D}_N(\hat{\rho}_N^*)\phi_0\\
&\quad + \tfrac{1}{N_1\hat{\sigma}_{v,N}^{*2}}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathbf{B}_N^{-1\prime}\mathbf{B}_N'(\hat{\rho}_N^*)\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbf{B}_N(\hat{\rho}_N^*)\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1)\\
&= \tfrac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbb{D}_N'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}_N\phi_0 + \tfrac{1}{N_1\sigma_{v0}^2}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1)\\
&= \tfrac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbb{D}_N'\mathcal{P}_2'\mathcal{P}_2\mathbb{D}_N\phi_0 + \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}] + o_p(1).
\end{aligned}
$$

We see that the bias term, $\frac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}}]$, involves only the common parameters that can be consistently estimated. Thus, a bias correction can easily be made. Define

$$
\mathrm{Bias}_{\lambda\lambda}^*(\delta) = \tfrac{1}{N_1}\mathtt{tr}[\mathcal{P}_2'(\delta)\mathcal{P}_2(\delta)\mathbb{P}_{\mathbb{D}}(\rho)]. \tag{D.3}
$$

This gives the bias matrix of $\Gamma_N^*(\hat{\theta}_N^*)$, which is a matrix of the same dimension as $\Gamma_N^*(\hat{\theta}_N^*)$, and has the sole non-zero element $\mathrm{Bias}_{\lambda\lambda}^*(\delta_0)$ corresponding to the $\Gamma_{\lambda\lambda}^*(\hat{\theta}_N^*)$ component. ■

Proof of Corollary 2.2.

Proof of (i). Note: $\mathbf{V} = \mathbf{B}_N(\mathbf{A}_N\mathbf{Y} - \eta)$, $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat{\rho}_N^*)\mathbf{B}_N(\hat{\rho}_N^*)[\mathbf{A}_N(\hat{\lambda}_N^*)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*]$ with respective elements $\{v_j\}, \{\tilde{v}_j\}$ and $\{\hat{v}_j\}$, and $\mathbb{Q}_{\mathbb{D}}$ has elements $\{q_{jh}\}$, $j, h = 1, \ldots, N$, where j and h are the combined indices for $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$.

Consistency of $\hat{\gamma}_N$. As $\hat{\sigma}_{v,N}^* - \sigma_{v0} = o_p(1)$ and $\hat{\rho}_N^* - \rho_0 = o_p(1)$, the denominators of $\hat{\gamma}_N$ and γ agree asymptotically. Thus, $\hat{\gamma}_N$ is consistent if $\frac{1}{N}\sum_{j=1}^N[\hat{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0$, or

$$
(a)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] \xrightarrow{p} 0, \text{ and } \quad (b)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) \xrightarrow{p} 0.
$$

To prove (a), note that $\tilde{v}_j = \sum_{h=1}^N q_{jh}v_h$. Thus, we have,

$$
\begin{aligned}
\tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] &= \tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{h=1}^N q_{jh}^3[v_h^3 - \mathrm{E}(v_h^3)] + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}v_l^2 v_m\\
&\quad + \tfrac{6}{N}\textstyle\sum_{j=1}^N\sum_{m=1}^N\sum_{\substack{l\neq m\\ l=1}}^N\sum_{\substack{h\neq m,l\\ h=1}}^N q_{jm}q_{jl}q_{jh}v_m v_l v_h \equiv K_1 + K_2 + K_3.
\end{aligned}
$$

First, consider K_1 term. By Lemma A.3, $\mathbb{Q}_{\mathbb{D}}$ is uniformly bounded in both row and column sums. This implies that the elements of $\mathbb{Q}_{\mathbb{D}}$ are uniformly bounded. Therefore, there exists a constant $\bar{q}$ such that $|q_{jh}| \leq \bar{q}$ for all j and h. Given these, we have $\sum_{j=1}^N q_{jh}^3 \leq \sum_{j=1}^N |q_{jh}|^3 \leq \bar{q}^2 \sum_{j=1}^N |q_{jh}| < \infty$. Also note $\{v_i\}$ are iid by Assumption A. Thus, Khinchine's weak law of large number (WLLN) (Feller, 1968, pp. 243-244) implies that K_1 converges to zero in probability as sample size increases.

For the other two terms, we have by switching the order of summations when needed,

$$
\begin{aligned}
K_2 &= \tfrac{3}{N}\sum_{j=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}(v_l^2-\sigma_v^2)v_m + \tfrac{3}{N}\sum_{j=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m,\\
&= \tfrac{3}{N}\sum_{m=1}^N (v_m^2-\sigma_v^2)(\sum_{j=1}^N \sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l) + \tfrac{3}{N}\sum_{m=1}^N v_m[\sum_{j=1}^N \sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2-\sigma_v^2)],\\
&\quad + \tfrac{3}{N}\sum_{m=1}^N \sum_{j=1}^N \sum_{\substack{l\neq m\\ l=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m,\\
K_3 &= \tfrac{18}{N}\sum_{m=1}^N v_m(\sum_{j=1}^N \sum_{l=1}^{m-1} \sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h) \equiv \tfrac{1}{N}\sum_{m=1}^N g_{4,m}.
\end{aligned}
$$

Therefore, we have $K_2 = \frac{1}{N}\sum_{m=1}^N (g_{1,m}+g_{2,m}+g_{3,m})$ and $K_3 = \frac{1}{N}\sum_{m=1}^N g_{4,m}$, where

$$
\begin{aligned}
g_{1,m} &= 3(v_m^2-\sigma_v^2)\sum_{j=1}^N \sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l,\\
g_{2,m} &= 3v_m \sum_{j=1}^N \sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2-\sigma_v^2),\\
g_{3,m} &= 3\sum_{j=1}^N \sum_{\substack{l\neq m\\ l=1}}^N q_{jl}^2 q_{jm}\sigma_v^2 v_m,\\
g_{4,m} &= v_m \sum_{j=1}^N \sum_{l=1}^{m-1} \sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h.
\end{aligned}
$$

Let $\{\mathcal{G}_m\}$ be the increasing sequence of σ-fields generated by $(v_1,\cdots,v_j, j=1,\cdots,m)$, $m=1,\cdots,N$. Then, $\mathrm{E}[(g_{1,m},g_{2,m},g_{3,m},g_{4,m})|\mathcal{G}_{m-1}]=0$; hence, $\{(g_{1,m},g_{2,m},g_{3,m},g_{4,m})',\mathcal{G}_m\}$ form a vector martingale difference (M.D.) sequence. As $\mathbb{Q}_{\mathbb{D}}$ is bounded in row and column sum norms, by Assumption A, it is easy to see that $\mathrm{E}|g_{s,m}|^{1+\epsilon} < \infty$, for $s=1,2,3,4$ and $\epsilon > 0$. Hence, $\{g_{1,m}\}$, $\{g_{2,m}\}$, $\{g_{3,m}\}$ and $\{g_{4,m}\}$ are uniformly integrable, and the WLLN of Davidson (1994, Theorem 19.7) applies to give $K_2 \xrightarrow{p} 0$ and $K_3 \xrightarrow{p} 0$.

To prove (b), using the notation $\tilde{\mathbf{V}}(\xi) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y}-\mathbf{X}\beta]$ in (2.5) where $\xi = (\beta',\delta')'$, we have $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}(\xi_0)$ and $\hat{\mathbf{V}} = \tilde{\mathbf{V}}(\hat{\xi}_N^*)$. Let $\mathbf{S}(\xi) = \frac{\partial}{\partial \xi'}\tilde{\mathbf{V}}(\xi)$, we have

$$
\mathbf{S}(\xi) = \{-\mathbb{X}(\rho),\ \ -\mathbb{Y}(\rho),\ \ [\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M}][\mathbf{A}_N(\lambda)\mathbf{Y}-\mathbf{X}\beta]\},
$$

where expressions of $\mathbb{Y}(\rho)$ and $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)$ are in (B.1) and (B.4), respectively. Let $s_j'(\xi)$ be the jth row of $\mathbf{S}(\xi)$. We have by the MVT, for each $j=1,2,\ldots,N$,

$$
\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_N^*) = \tilde{v}_j(\xi_0) + s_j'(\bar{\xi})(\hat{\xi}_N^* - \xi_0) = \tilde{v}_j + \psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|), \tag{D.4}
$$

where $\bar{\xi}$ lies between $\hat{\xi}_N^*$ and ξ_0, and $\psi_j' = \mathrm{plim}_{N\to\infty} s_j'(\bar{\xi})$, which is easily shown to be $O_p(1)$ as follow. Consider the first k (the number of regressors) elements of ψ_j' first. They are the limits of the jth row of $-\mathbb{X}(\bar{\rho})$, which are just the jth row of $-\mathbb{X}$ because $\bar{\rho} \xrightarrow{p} \rho_0$, implied by $\hat{\rho}_N^* - \rho_0 = o_p(1)$. Hence, we conclude that the first k elements of ψ_j' are $O(1)$, for each $j = 1, 2, \ldots, N$. For the remaining two elements in each ψ_j', they are the limits of elements from the last two columns of $\mathbf{S}(\bar{\xi})$. It is easy to see the limits of the last two columns of $\mathbf{S}(\bar{\xi})$ are just $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0]$. Using $\mathbf{Y} = \mathbf{A}_N^{-1}\eta + \mathbf{C}_N^{-1}\mathbf{V}$, we have $-\mathbb{Y} = \mathcal{P}_2\mathbf{B}_N\eta + \mathcal{P}_2\mathbf{V}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0] = [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0 + [\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_N^{-1}\mathbf{V}$. By Lemma A.1, we have the elements of $\mathcal{P}_2\mathbf{B}_N\eta$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{D}\phi_0$ are uniformly bounded, and $\mathcal{P}_2$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}]\mathbf{B}_N^{-1}$ are uniformly bounded in both row and column sum norms. Hence, it is easy to see each element of $-\mathbb{Y}$ and $[\dot{\mathbb{Q}}_{\mathbb{D}}\mathbf{B}_N - \mathbb{Q}_{\mathbb{D}}\mathbf{M}][\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0]$ are $O_p(1)$, i.e., the last two elements in ψ_j' are also $O_p(1)$, for each $j = 1, 2, \ldots, N$.

As $\tilde{v}_j = O_p(1)$, $\psi_j' = O_p(1)$ and $\hat{\xi}_N^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have by (D.4), $\hat{v}_j^3 = \tilde{v}_j^3 + 3\tilde{v}_j^2\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)$. It follows that

$$\begin{aligned}
\tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) &= \tfrac{3}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^2\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|) \\
&= \tfrac{3\sigma_v^2}{N}\textstyle\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j')(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|) = o_p(1),
\end{aligned}$$

as $\frac{1}{N}\sum_{j=1}^N(\sum_{k=1}^N q_{jk}^2\psi_j') = (\sum_{k=1}^N q_{jk}^2)\frac{1}{N}(\sum_{j=1}^N \psi_j') = O(1)$.

Consistency of $\hat{\kappa}_N$. As $\hat{\sigma}_{v,N}^* - \sigma_{v0} = o_p(1)$ and $\hat{\rho}_N^* - \rho_0 = o_p(1)$, the result follows if $\frac{1}{N}\sum_{j=1}^N[\hat{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0$. This amounts to show that

$$(c)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0 \ \text{ and } \ (d)\ \tfrac{1}{N}\sum_{j=1}^N(\hat{v}_j^4 - \tilde{v}_j^4) \xrightarrow{p} 0.$$

To prove (c), we have

$$\begin{aligned}
&\tfrac{1}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^4 - \tfrac{1}{N}\sum_{j=1}^N \mathrm{E}(\tilde{v}_j^4) \\
=&\tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{h=1}^N q_{jh}^4[v_h^4 - \mathrm{E}(v_h^4)] + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2q_{jm}^2(v_l^2v_m^2 - \sigma_v^4) \\
&+ \tfrac{4}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^3q_{jm}v_l^3v_m + \tfrac{6}{N}\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N\sum_{\substack{h\neq m,l\\ m=1}}^N q_{jl}^2q_{jm}q_{jh}v_l^2v_mv_h \\
&+ \tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^N\sum_{\substack{m\neq l\\ m=1}}^N\sum_{\substack{h\neq m,l\\ m=1}}^N\sum_{\substack{p\neq m,l,h\\ m=1}}^N q_{jl}q_{jm}q_{jh}q_{jp}v_lv_mv_hv_p \equiv \sum_{r=1}^5 R_r.
\end{aligned}$$

By using WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays as in the proof of (a), we have $R_r = o_p(1)$ for $r = 1, 3, 4, 5$. For R_2, noting that $v_l^2v_m^2 - \sigma_v^4 = (v_l^2 - \sigma_v^2)(v_m^2 - \sigma_v^2) + \sigma_v^2(v_m^2 - \sigma_v^2) + \sigma_v^2(v_l^2 - \sigma_v^2)$, we have

$$\begin{aligned}
R_2 = &\tfrac{6}{N}\textstyle\sum_{l=1}^N(v_l^2 - \sigma_v^2)[\sum_{j=1}^N\sum_{m=1}^{l-1} q_{jl}^2q_{jm}^2(v_m^2 - \sigma_v^2)] \\
&+ \tfrac{6}{N}\textstyle\sum_{l=1}^N[\sum_{j=1}^N\sum_{\substack{m\neq l\\ m=1}}^N q_{jl}^2q_{jm}^2\sigma_v^2(v_l^2 - \sigma_v^2)] \equiv \tfrac{6}{N}\sum_{l=1}^N(f_l + f_{2,l}).
\end{aligned}$$

Since $\mathrm{E}[f_l|\mathcal{G}_{l-1}] = 0$ and $\{f_{2,l}\}$ are independent, it is easy to see they both form an M.D. sequence. In addition, it is easily seen that $\mathrm{E}|f_{s,l}|^{1+\epsilon} < \infty$, for $s = 1, 2$ and $\epsilon > 0$, so that $\{f_l\}$ and $\{f_{2,l}\}$ are uniformly integrable. Therefore, the WLLN of Davidson (1994, Theorem 19.7) also implies that $\frac{6}{N}\sum_{l=1}^{N} f_l = o_p(1)$ and $\frac{6}{N}\sum_{l=1}^{N} f_{2,l} = o_p(1)$.

To prove (d), we have by (D.4) $\hat{v}_j^4 = \tilde{v}_j^4 + 4\tilde{v}_j^3\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)$. It follows that

$$\begin{aligned}\tfrac{1}{N}\textstyle\sum_{j=1}^{N}(\hat{v}_j^4 - \tilde{v}_j^4) &= \tfrac{4}{N}\textstyle\sum_{j=1}^{N}\tilde{v}_j^3\psi_j'(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|)\\ &= \tfrac{4\sigma_v^3\gamma}{N}\textstyle\sum_{j=1}^{N}(\sum_{k=1}^{N} q_{jk}^3\psi_j')(\hat{\xi}_N^* - \xi_0) + o_p(\|\hat{\xi}_N^* - \xi_0\|) = o_p(1).\end{aligned}$$

Proof of (ii). The consistency of $\widehat{\Sigma}_N^*$ to $\Sigma_N^*(\theta_0)$ can be shown similarly as what we do in the proof of Theorem 2.2 for results (b) and (c). For $\widehat{\Gamma}_N^* - \Gamma_N^*(\theta_0) \xrightarrow{p} 0$, we only need to show that $\mathrm{Bias}^*(\hat{\delta}_N^*) - \mathrm{Bias}^*(\delta_0) = o_p(1)$, based on Corollary 2.1. That is to show

$$\tfrac{1}{N_1}\{\mathtt{tr}[\mathcal{P}_2'(\hat{\delta}_N^*)\mathcal{P}_2(\hat{\delta}_N^*)\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^*)] - \mathtt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})\} = o_p(1),$$

which can be easily proved by using the MVT as we do for $\frac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar{\delta}) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]$ in the proof of Theorem 2.2 (b). ∎

Proof of Theorem 3.2. Applying the MVT on each row of $S_N^\diamond(\hat{\xi}_N^\diamond)$, we have,

$$0 = \tfrac{1}{\sqrt{N_1}}S_N^\diamond(\hat{\xi}_N^\diamond) = \tfrac{1}{\sqrt{N_1}}S_N^\diamond(\xi_0) + \left[\tfrac{1}{N_1}\tfrac{\partial}{\partial\xi'}S_N^\diamond(\xi)\Big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}}\right]\sqrt{N_1}(\hat{\xi}_N^\diamond - \xi_0),$$

where $\{\bar{\xi}_r\}$ are on the line segment between $\hat{\xi}_N^\diamond$ and ξ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N_1}}S_N^\diamond(\xi_0) \xrightarrow{D} N[0,\ \lim_{N\to\infty}\Gamma_N^\diamond(\xi_0)]$,

(b) $\frac{1}{N_1}\big[\frac{\partial}{\partial\xi'}S_N^\diamond(\xi)\big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial\xi'}S_N^\diamond(\xi_0)\big] = o_p(1)$, and

(c) $\frac{1}{N_1}\big[\frac{\partial}{\partial\xi'}S_N^\diamond(\xi_0) - \mathrm{E}(\frac{\partial}{\partial\xi'}S_N^\diamond(\xi_0))\big] = o_p(1)$.

Proof of (a). From (3.6), we see that the elements of $S_N^\diamond(\xi_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+2)\times 1$ vector of constants a, $a'S_N^\diamond(\xi_0)$ has form:

$$a'S_N^\diamond(\xi_0) = b_N'\mathbf{V} + \mathbf{V}'\Phi_N\mathbf{V} - \sigma_v^2\mathtt{tr}(\Phi_N),$$

for suitably defined non-stochastic vector b_N and matrix Φ_N. Again, by Assumptions A-F it is easy to verify that b_N and matrix Φ_N satisfy the conditions of the CLT for LQ form of Kelejian and Prucha (2001), and hence the asymptotic normality of $\frac{1}{\sqrt{N_1}}a'S_N^\diamond(\xi_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}}S_N^\diamond(\xi_0) \xrightarrow{D} N[0, \lim_{N\to\infty}\Gamma_N^\diamond(\theta_0)]$, where $\Gamma_N^\diamond(\theta_0)$ is given in (3.7).

Proof of (b). The Hessian matrix $H_N^\diamond(\xi) = \frac{\partial}{\partial\xi'}S_N^\diamond(\xi)$ is given in (C.1). As $\bar{\mathbb{F}}_{N\lambda}'(\delta_0)$,

$\bar{\mathbb{F}}'_{N\rho}(\delta_0)$ and $\bar{\mathbb{G}}_{N\rho}(\rho_0)$ are diagonal matrices with uniformly bounded elements, it is easy to see that $\frac{1}{N_1}H_N^{\diamond}(\xi_0) = O_p(1)$ by Lemma A.6, and hence, $\frac{1}{N_1}H_N^{\diamond}(\bar{\xi}) = O_p(1)$. Here again for ease of exposition we simply use $H_N^{\diamond}(\bar{\xi})$ to denote $\frac{\partial}{\partial \xi'}S_N^{\diamond}(\xi)\big|_{\xi=\bar{\xi}_r \text{ in } r\text{th row}}$. As $H_N^{\diamond}(\bar{\xi})$ is linear or quadratic in $\bar{\beta}$ and nonlinear in $\bar{\delta}$, we have by applying the MVT on the $\bar{\delta}$-components:

$$\tfrac{1}{N_1}H_N^{\diamond}(\bar{\xi}) - \tfrac{1}{N_1}H_N^{\diamond}(\xi_0) = \tfrac{1}{N_1}\tfrac{\partial}{\partial \delta'}H_N^{\diamond}(\bar{\beta},\dot{\delta})(\bar{\delta}-\delta_0) + \tfrac{1}{N_1}[H_N^{\diamond}(\bar{\beta},\delta_0) - H_N^{\diamond}(\theta_0)].$$

Similar to the proof of Theorem 2.2 (b), we show that $\frac{1}{N_1}\frac{\partial}{\partial \delta'}H_N^{\diamond}(\bar{\beta},\dot{\delta}) = O_p(1)$. The second term is seen to contain elements either linear or quadratic in $\bar{\beta}-\beta_0$ with the matrices in the linear or quadratic terms being $O_p(1)$. Hence, the desired result follows as $\bar{\xi}-\xi_0 = o_p(1)$.

Proof of (c). Since $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, all components of $H_N^{\diamond}(\xi_0)$ are linear or quadratic in $\mathbf{V}$. Thus, under the assumptions of the theorem the result **(c)** is proved using Lemma A.6. We provide details of the proof using the most complicate term, $H_{\rho\rho}^{\diamond}(\xi_0)$. Let $\Xi_N = -\mathbf{G}'_N\bar{\mathbb{G}}_N + \bar{\mathbb{G}}_{N\rho} + \bar{\mathbb{G}}_N\mathbb{G}_N$. By Lemma A.1, it is easy to see that Ξ_N is uniformly bounded in both row and column sums in absolute value. Hence, we have

$$\begin{aligned}
&\tfrac{1}{N_1}[H_{\rho_0\rho_0}^{\diamond}(\xi_0) - \mathrm{E}(H_{\rho_0\rho_0}^{\diamond}(\xi_0))]\\
&= \tfrac{1}{N_1}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] - \tfrac{1}{N_1}(\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0)'\mathbf{B}'_N\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}\\
&\quad + \tfrac{1}{N_1}\mathrm{E}[(\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0)'\mathbf{B}'_N\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}]\\
&= \tfrac{1}{N_1}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})] - \tfrac{1}{N_1}[\phi_0'\mathbb{D}'_N\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\phi_0'\mathbb{D}'_N\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V})]\\
&\quad - \tfrac{1}{N_1}[\mathbf{V}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathrm{E}(\mathbf{V}'\Xi_N\mathbb{Q}_{\mathbb{D}}\mathbf{V})] \;=\; o_p(1).
\end{aligned}$$

The proofs for the other terms are done in a similar manner, and the details are omitted. ■

Proof of Corollary 3.1: Just like the homoskedasiticiy case, plugging $\hat{\phi}_N^{\diamond}$ in $\Gamma_N^{\diamond}(\xi)$ induces a bias for terms quadratic in ϕ, and a bias correction is necessary. From (3.7), we see that the terms of $\Gamma_N^{\diamond}(\xi)$ that are quadratic in ϕ are the (λ,ρ) terms and are of the form: $\phi'\mathbb{D}'_N(\rho)\mathbb{L}'_a(\delta)\mathbf{H}\mathbb{L}_b(\delta)\mathbb{D}_N(\rho)\phi,\ a,b = \lambda,\rho$, recalling $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$ and $\mathbb{D}(\rho)\mathbf{B}_N(\rho)\mathbf{D}$.

By applying the MVT on $\hat{\rho}_N^{\diamond}$-variable in the key quantity $\mathbf{D}\hat{\phi}_N^{\diamond}$, we have after some algebra,

$$\begin{aligned}
\mathbf{D}\hat{\phi}_N^{\diamond} &= \mathbf{D}\phi_0 + \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} - \mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{B}_N[\mathbf{W}\mathbf{Y}(\hat{\lambda}_N^{\diamond}-\lambda_0) + \mathbf{X}(\hat{\beta}_N^{\diamond}-\beta_0)]\\
&\quad - \mathbb{R}_N(\dot{\rho})[\mathbf{A}_N(\hat{\lambda}_N^{\diamond})\mathbf{Y} - \mathbf{X}\hat{\beta}_N^{\diamond}](\hat{\rho}_N^{\diamond}-\rho_0),
\end{aligned}$$

where $\dot{\rho}$ lies between $\hat{\rho}_N^{\diamond}$ and ρ_0. Plugging $\mathbf{D}\hat{\phi}_N^{\diamond}$ and other parameter estimates in these quadratic terms, we have,

$$
\begin{aligned}
&\tfrac{1}{N_1}\hat{\phi}_N^{\diamond\prime}\mathbb{D}_N'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{D}_N(\hat{\rho}_N^{\diamond})\hat{\phi}_N^{\diamond} \\
&= \tfrac{1}{N_1}\phi_0'\mathbb{D}_N'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{D}_N(\hat{\rho}_N^{\diamond})\phi_0 \\
&\quad + \tfrac{1}{N_1}\mathbf{V}'\mathbb{P}_{\mathbb{D}}\mathbf{B}_N^{-1\prime}\mathbf{B}_N'(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\mathbf{H}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbf{B}_N(\hat{\rho}_N^{\diamond})\mathbf{B}_N^{-1}\mathbb{P}_{\mathbb{D}}\mathbf{V} + o_p(1) \\
&= \tfrac{1}{N_1}\phi_0'\mathbb{D}_N'\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{D}_N\phi_0 + \tfrac{1}{N_1}\mathtt{tr}\big[\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\big] + o_p(1),
\end{aligned}
$$

Define

$$
\mathrm{Bias}_{ab}^{\diamond}(\delta,\mathbf{H}) = \tfrac{1}{N_1}\mathtt{tr}\big[\mathbf{H}\mathbb{P}_{\mathbb{D}}(\rho)\mathbb{L}_a'(\delta)\mathbf{H}\mathbb{L}_b(\delta)\mathbb{P}_{\mathbb{D}}(\rho)\big],
$$

for $a, b = \lambda, \rho$. Hence, the bias matrix for $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond})$ can be written as

$$
\mathrm{Bias}_{\phi}^{\diamond}(\delta_0,\mathbf{H}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathrm{Bias}_{\lambda\lambda}^{\diamond}(\delta_0,\mathbf{H}) & \mathrm{Bias}_{\lambda\rho}^{\diamond}(\delta_0,\mathbf{H}) \\ 0 & \mathrm{Bias}_{\rho\lambda}^{\diamond}(\delta_0,\mathbf{H}) & \mathrm{Bias}_{\rho\rho}^{\diamond}(\delta_0,\mathbf{H}) \end{bmatrix},
$$

leading to the result of Corollary 3.1. ■

Proof of Lemma 3.1: Using $\tilde{\mathbf{V}}(\xi) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ defined in (2.5), let $\tilde{\mathbf{V}} = \tilde{\mathbf{V}}(\xi_0)$ and $\hat{\mathbf{V}} = \tilde{\mathbf{V}}(\hat{\xi}_N^{\diamond})$ and denote their elements by $\{\tilde{v}_j\}$ and $\{\hat{v}_j\}$, respectively. Following (D.4), we have $\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_N^{\diamond}) = \tilde{v}_j + \psi_j'(\hat{\xi}_N^{\diamond} - \xi_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|)$, and in vector form,

$$
\hat{\mathbf{V}} = \tilde{\mathbf{V}} + \Psi_N(\hat{\xi}_N^{\diamond} - \xi_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|),
$$

where $\Psi_N = (\psi_1, \psi_2, \ldots, \psi_N)'$, with ψ_j being defined below (D.4).

Define $\dot{\Pi}_N(\rho) = \frac{\partial}{\partial\rho}\Pi_N(\rho) = -2\Pi_N(\rho)[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]\Pi_N(\rho)$. It is easy to see that $\|\dot{\Pi}_N(\rho)\|_1$ and $\|\dot{\Pi}_N(\rho)\|_\infty$ are bounded in a neighborhood of ρ_0. Let Π_{jh} and $\dot{\Pi}_{jh}$ be the respective elements of Π_N and $\dot{\Pi}_N$. Hence, we have by the MVT, for each $j, h = 1, 2, \ldots, N$, $\Pi_{jh}(\hat{\rho}_N^{\diamond}) = \Pi_{jh} + \dot{\Pi}_{jh}(\bar{\rho})(\hat{\rho}_N^{\diamond} - \rho_0) = \Pi_{jh} + \dot{\Pi}_{jh}(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\rho}_N^{\diamond} - \rho_0\|)$, where $\bar{\rho}$ lies between $\hat{\rho}_N^{\diamond}$ and ρ_0. In matrix form, we have

$$
\Pi_N(\hat{\rho}_N^{\diamond}) = \Pi_N + \dot{\Pi}_N(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\rho}_N^{\diamond} - \rho_0\|).
$$

Define $\hat{h} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)' = \Pi_N(\hat{\rho}_N^{\diamond})(\hat{\mathbf{V}} \odot \hat{\mathbf{V}})$ and $\tilde{h} = \Pi_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})$. As the elements of $\tilde{\mathbf{V}}$ are $O_p(1)$, rows of Ψ_N are $O_p(1)$, elements of Π_N and $\dot{\Pi}_N$ are $O(1)$, and $\hat{\xi}_N^{*} - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have,

$$
\hat{h} = \tilde{h} + 2\Pi_N(\tilde{\mathbf{V}} \odot \Psi_N(\hat{\xi}_N^{\diamond} - \xi_0)) + \dot{\Pi}_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|). \qquad \text{(D.5)}
$$

Proof of (i). Let $c_N = (c_{11}, \cdots, c_{NN})'$ and $h = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$. We have,

$$\tfrac{1}{N}[\mathtt{tr}(\hat{\mathbf{H}}C_N) - \mathtt{tr}(\mathbf{H}C_N)] = \tfrac{1}{N}c_N'(\hat{h} - h) = \tfrac{1}{N}c_N'(\hat{h} - \tilde{h}) + \tfrac{1}{N}c_N'(\tilde{h} - h).$$

The result follows if both terms above are $o_p(1)$. For the first term, we have, using (D.5),

$$\begin{aligned}
&\tfrac{1}{N}\bar{c}_N'(\hat{h} - \tilde{h}) = \tfrac{2}{N}\bar{c}_N'\Pi_N(\tilde{\mathbf{V}} \odot \Psi_N(\hat{\xi}_N^{\diamond} - \xi_0)) + \tfrac{1}{N}\bar{c}_N'\dot{\Pi}_N(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}})(\hat{\rho}_N^{\diamond} - \rho_0) + o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|) \\
=&\tfrac{2}{N}\textstyle\sum_{j=1}^N c_{jj}(\sum_{h=1}^N \Pi_{jh}\tilde{v}_h\psi_h')(\hat{\xi}_N^{\diamond} - \xi_0) + \tfrac{1}{N}\sum_{j=1}^N c_{jj}(\sum_{h=1}^N \dot{\Pi}_{jh}\sum_{k=1}^N q_{hk}^2\sigma_k^2)(\hat{\rho}_N^{\diamond} - \rho_0) \\
&+ o_p(\|\hat{\xi}_N^{\diamond} - \xi_0\|) = o_p(1).
\end{aligned}$$

For the second term, we have after some algebra,

$$\tilde{h} = \Pi_N[(\mathbb{Q}_{\mathbb{D}} \odot \mathbb{Q}_{\mathbb{D}})(\mathbf{V} \odot \mathbf{V}) + \zeta] = \mathbf{V} \odot \mathbf{V} + \Pi_N\varepsilon, \tag{D.6}$$

where ε is an $N \times 1$ vector with j-th element $\varepsilon_j = \sum_{k=1}^N v_k\zeta_{jk}$, where $\zeta_{jk} = 2q_{jk}\sum_{l=1}^{k-1} q_{jl}v_l, k \geq 2$, and $\zeta_{j1} = 0$. As ζ_{jk} is $(v_1, \ldots v_{k-1})$-measurable, $\{v_k\zeta_{jk}\}$ form an M.D. sequence. Thus, each ε_j is a sum of M.D.s. Hence, we have

$$\tfrac{1}{N}\bar{c}_N'(\tilde{h} - h) = \tfrac{1}{N}\bar{c}_N'(\mathbf{V} \odot \mathbf{V} - h) + \tfrac{1}{N}\bar{c}_N'\Pi_N\zeta = o_p(1),$$

where $\frac{1}{N}\bar{c}_N'(\mathbf{V} \odot \mathbf{V} - h) = o_p(1)$ by Lemma A.6(v) and $\frac{1}{N}\bar{c}_N'\Pi_N\zeta = o_p(1)$ by WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays.

Proof of (ii). Note that $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = h'(A_N \odot B_N)h$. We have,

$$\begin{aligned}
&\tfrac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}A_N\widehat{\mathbf{H}}B_N) - \tfrac{1}{N}\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = \tfrac{1}{N}\hat{h}'(A_N \odot B_N)\hat{h} - \tfrac{1}{N}h'(A_N \odot B_N)h \\
=&\tfrac{1}{N}(\hat{h}'(A_N \odot B_N)\hat{h} - \tilde{h}'(A_N \odot B_N)\tilde{h}) + \tfrac{1}{N}(\tilde{h}'(A_N \odot B_N)\tilde{h} - h'(A_N \odot B_N)h).
\end{aligned} \tag{D.7}$$

The first term of (D.7) can be written as

$$\tfrac{1}{N}(\hat{h}'(A_N \odot B_N)\hat{h} - \tilde{h}'(A_N \odot B_N)\tilde{h}) = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3,$$

where $\mathcal{T}_1 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)(\hat{h} - \tilde{h})$, $\mathcal{T}_2 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)\tilde{h}$, and $\mathcal{T}_3 = \frac{1}{N}(\hat{h} - \tilde{h})'(A_N \odot B_N)'\tilde{h}$. Note that A_N and B_N are uniformly bounded in both row and column sum norms, $A_N \odot B_N$ is also uniformly bounded in both row and column sum norms. Hence, using (D.5), $\tilde{\mathbf{V}} = O_p(1)$, $\Psi_N = O_p(1)$ and $\hat{\xi}_N^* - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we can easily show that $\mathcal{T}_r = o_p(1)$, for $r = 1, 2, 3$, as we show $\frac{1}{N}\bar{c}_N'(\hat{h} - \tilde{h}) = o_p(1)$ in the proof of (i). Thus, the first term in (D.7) is $o_p(1)$.

For the second term in (D.7), we have similarly to the first term,

$$\tfrac{1}{N}(\tilde{h}'(A_N \odot B_N)\tilde{h} - h'(A_N \odot B_N)h) = \mathcal{T}_4 + \mathcal{T}_5 + \mathcal{T}_6,$$

where $\mathcal{T}_4 = \frac{1}{N}(\tilde{h}-h)'(A_N \odot B_N)(\tilde{h}-h)$, $\mathcal{T}_5 = \frac{1}{N}(\tilde{h}-h)'(A_N \odot B_N)h$ and $\mathcal{T}_6 = \frac{1}{N}(\tilde{h}-h)'(A_N \odot B_N)'h$. For the $\mathcal{T}_5$ term, we have by (D.6),

$$\mathcal{T}_5 = \tfrac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)h + \tfrac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)h = o_p(1),$$

by Lemma A.6(v) and WLLN for M.D. arrays of Davidson (1994, Theorem 19.7). The $\mathcal{T}_6$ term is similar to $\mathcal{T}_5$ and the result follows, i.e., $\mathcal{T}_6 = o_p(1)$.

Thus, it is left to study the limit of $\mathcal{T}_4$. Again, by (D.6) we have,

$$\begin{aligned}\mathcal{T}_4 &= \tfrac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)\Pi_N\varepsilon + \tfrac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)'\Pi_N\varepsilon \\ &\quad + \tfrac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)(\mathbf{V} \odot \mathbf{V} - h) + \tfrac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon \\ &\equiv \mathcal{T}_{4a} + \mathcal{T}_{4b} + \mathcal{T}_{4c} + \mathcal{T}_{4d}.\end{aligned} \tag{D.8}$$

Consider first the term $\mathcal{T}_{4a}$. Denote $\Omega = (A_N \odot B_N)\Pi_N$ with elements $\{\omega_{jk}\}$. We have,

$$\begin{aligned}\mathcal{T}_{4a} &= \tfrac{1}{N}\textstyle\sum_{j=1}^N \sum_{k=1}^N \omega_{jk}\varepsilon_j(v_k^2 - \sigma_k^2) \\ &= \tfrac{1}{N}\textstyle\sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \sum_{\substack{m\neq l\\ m=1}}^N \omega_{jk}q_{jl}q_{jm}(v_k^2 - \sigma_k^2)v_l v_m \\ &= \tfrac{1}{N}\textstyle\sum_{k=1}^N ((v_k^2 - \sigma_k^2)\sum_{j=1}^N \sum_{l=1}^{k-1} \sum_{\substack{m\neq l\\ m=1}}^{k-1} \omega_{jk}q_{jl}q_{jm}v_l v_m) \\ &\quad + \tfrac{2}{N}\textstyle\sum_{l=1}^N (v_l \sum_{j=1}^N \sum_{k=1}^{l-1} \sum_{\substack{m\neq k\\ m=1}}^{l-1} \omega_{jk}q_{jl}q_{jm}v_m(v_k^2 - \sigma_k^2)) \\ &\quad + \tfrac{2}{N}\textstyle\sum_{k=1}^N ((v_k^3 - \mathrm{E}v_k^3)\sum_{j=1}^N \sum_{m=1}^{k-1} \omega_{jk}q_{jk}q_{jm}v_m) \\ &\quad + \tfrac{2}{N}\textstyle\sum_{m=1}^N (v_m \sum_{j=1}^N \sum_{k=1}^{m-1} \omega_{jk}q_{jk}q_{jm}(v_k^3 - \mathrm{E}v_k^3)) \\ &\quad + \tfrac{2}{N}\textstyle\sum_{m=1}^N (v_m \sum_{j=1}^N \sum_{\substack{k\neq m\\ k=1}}^N \omega_{jk}q_{jk}q_{jm}(\mathrm{E}v_k^3 - \sigma_k^2)),\end{aligned}$$

which is seen to be the average of M.D. sequence and thus is $o_p(1)$ by Theorem 19.7 of Davidson (1994). Similarly, we show that $\mathcal{T}_{4b} = \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)'\Pi_N\varepsilon = o_p(1)$.

For the term $\mathcal{T}_{4c}$, as $\mathrm{E}(\mathbf{V} \odot \mathbf{V}) = h$, we have $\mathrm{E}(\mathcal{T}_{4c}) = \frac{1}{N}\mathtt{tr}((A_N \odot B_N)\mathrm{Var}(\mathbf{V} \odot \mathbf{V})) = 0$. Thus, Lemma A.6($iv$) implies that $\mathcal{T}_{4c} = \frac{1}{N}(\mathbf{V} \odot \mathbf{V} - h)'(A_N \odot B_N)(\mathbf{V} \odot \mathbf{V} - h) \xrightarrow{p} 0$.

Now, for the last term of (D.8), $\mathcal{T}_{4d} = \frac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon$, we have by taking the advantage that each element of ε is a sum of an M.D. sequence,

$$\mathrm{E}(\varepsilon\varepsilon') = 2(\mathbb{Q}_\mathbb{D}\mathbf{H}\mathbb{Q}_\mathbb{D}) \odot (\mathbb{Q}_\mathbb{D}\mathbf{H}\mathbb{Q}_\mathbb{D}) - 2(\mathbb{Q}_\mathbb{D} \odot \mathbb{Q}_\mathbb{D})\mathbf{H}\mathbf{H}(\mathbb{Q}_\mathbb{D} \odot \mathbb{Q}_\mathbb{D}). \tag{D.9}$$

This gives,

$$\mathrm{E}(\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon) = 2\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) - 2\mathtt{tr}((A_N \odot B_N)\mathbf{H}^2), \quad \text{(D.10)}$$
$$= 2\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N),$$

where $\Lambda(\mathbf{H}) = (\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}}) \odot (\mathbb{Q}_{\mathbb{D}}\mathbf{H}\mathbb{Q}_{\mathbb{D}})$, and the last equation takes use of the fact that the diagonal elements of A_N and B_N are zero.

Finally, to show that $\mathcal{T}_{4d} - \mathrm{E}(\mathcal{T}_{4d}) = o_p(1)$, denote $\chi_N = \Pi_N(A_N \odot B_N)\Pi_N$ with elements $\{\chi_{jk}\}$. It is easy to show that $\{\chi_{jk}\}$ are uniformly bounded, and let $|\chi_{lm}| \leq \bar{\chi} < \infty$. We have,

$$\begin{aligned}
&\mathrm{Var}(\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon) \\
&= 8\textstyle\sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \sum_{m=1}^N \sum_{h=1}^N \sum_{\substack{p\neq h\\p=1}}^N \sum_{s=1}^N \sum_{\substack{r\neq s\\r=1}}^N \\
&\qquad \chi_{jk}\chi_{lm}q_{jh}q_{jp}q_{lh}q_{lp}q_{ks}q_{kr}q_{ms}q_{mr}\mathrm{E}(v_h^2v_p^2v_s^2v_r^2) \\
&\leq 8\bar{q}^2\bar{\chi}c\textstyle\sum_{m=1}^N(\sum_{j=1}^N|\chi_{jk}|)(\sum_{k=1}^N|q_{kr}|)(\sum_{l=1}^N|q_{lp}|)(\sum_{h=1}^N|q_{lh}|) \\
&\qquad (\textstyle\sum_{p=1}^N|q_{jp}|)(\sum_{s=1}^N|q_{ms}|)(\sum_{r=1}^N|q_{mr}|) \\
&= O(N),
\end{aligned}$$

where the inequality holds because $\mathrm{E}(v_h^2v_p^2v_s^2v_r^2)$ equals either $\mathrm{E}(v_h^2v_s^2)\mathrm{E}(v_p^2v_r^2)$ or $\mathrm{E}(v_h^2v_r^2)\mathrm{E}(v_p^2v_s^2)$ since $h \neq p$ and $s \neq r$, and either of them is less than a constant $c < \infty$, e.g., $\mathrm{E}(v_h^2v_r^2) \leq \mathrm{E}^{\frac{1}{2}}(v_h^4)\mathrm{E}^{\frac{1}{2}}(v_r^4) \leq c$. Therefore, by Chebyshev's inequality,

$$\begin{aligned}
&P(\tfrac{1}{N}|\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon - \mathrm{E}(\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon)| \geq M) \\
&\leq \tfrac{1}{M^2}\tfrac{1}{N^2}\mathrm{Var}(\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon) = o(1).
\end{aligned}$$

It follows that $\frac{1}{N}\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon - \frac{1}{N}\mathrm{E}(\varepsilon'\Pi_N(A_N \odot B_N)\Pi_N\varepsilon) \xrightarrow{p} 0$. Therefore, we have shown that $\mathcal{T}_4 = \frac{2}{N}\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) + o_p(1)$. It follows that

$$\begin{aligned}
&\tfrac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}A_N\widehat{\mathbf{H}}B_N^\circ) - \tfrac{1}{N}\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N^\circ) = \textstyle\sum_{r=1}^6 \mathcal{T}_r \\
&= \tfrac{2}{N}\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) + o_p(1),
\end{aligned}$$

completing the proof of Lemma 3.1. ■

Proof of Corollary 3.2: The consistency of $\widehat{\Sigma}_N^\diamond$ to $\Sigma_N^\diamond(\xi_0)$ is implied by results (b) and (c) in the proof of Theorem 3.2. To show $\widehat{\Gamma}_N^\diamond - \Gamma_N^\diamond(\xi_0) \xrightarrow{p} 0$, we argue as follows:

(a) The transition from $\Gamma_N^\diamond(\xi_0, \phi_0, \mathbf{H})$ to $\Gamma_N^\diamond(\hat{\xi}_N^\diamond, \phi_0, \mathbf{H})$ does not incur cost asymptotically;

(b) The cost of transition from $\Gamma_N^\diamond(\hat{\xi}_N^\diamond, \phi_0, \mathbf{H})$ to $\Gamma_N^\diamond(\hat{\xi}_N^\diamond, \hat{\phi}_N^\diamond, \mathbf{H})$ is captured by $\mathrm{Bias}_\phi^\diamond(\hat{\delta}_N^\diamond, \mathbf{H})$;

(c) The effect of replacing $\mathbf{H}$ in $\frac{1}{N_1}\mathtt{tr}(\mathbf{H}\mathbb{L}_a\mathbf{H}\mathbb{L}_b^{\circ}), a, b = \lambda, \rho$, is captured by $\frac{2}{N_1}\mathtt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^{\circ})\Pi_N\Lambda(\mathbf{H})\Pi_N), a, b = \lambda, \rho$;

(d) It is left to show that the cost of transition from $\text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \mathbf{H})$ to $\text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \hat{\mathbf{H}})$ is captured by $-\frac{2}{N_1}\mathtt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N), a, b = \lambda, \rho$.

The non-zero entries in $\text{Bias}_{\phi}^{\diamond}(\delta_0, \mathbf{H})$ are of the form $\frac{1}{N_1}\mathtt{tr}(\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}})$, for $a, b = \lambda, \rho$, as given in Corollary 3.1. Applying result (D.10) with $A_N = \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'$ and $B_N = \mathbb{L}_b\mathbb{P}_{\mathbb{D}}$, we have,

$$
\begin{aligned}
&\tfrac{1}{N_1}\mathtt{tr}\big[\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^{\diamond})\mathbb{L}_a'(\hat{\delta}_N^{\diamond})\hat{\mathbf{H}}\mathbb{L}_b(\hat{\delta}_N^{\diamond})\mathbb{P}_{\mathbb{D}}(\hat{\rho}_N^{\diamond})\hat{\mathbf{H}} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\mathbf{H}\big] \\
=&\tfrac{1}{N_1}\mathtt{tr}\big[\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\hat{\mathbf{H}}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\hat{\mathbf{H}} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}}\mathbf{H}\big] + o_p(1) \;\; \text{(by the MVT)} \\
=&\tfrac{2}{N_1}\mathtt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N) + \tfrac{1}{N}\mathtt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\mathbf{H}^2) + o_p(1), \\
=&\tfrac{2}{N_1}\mathtt{tr}((\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N) + o_p(1),
\end{aligned}
$$

for $a, b = \lambda, \rho$. Although the diagonal elements of $\mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}}$ may not be zero uniformly, their magnitudes are typically small so that the second term of the second last equation is negligible.[11] Then, it follows that $\widehat{\Sigma}_N^{\diamond-1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond-1} - \Sigma_N^{\diamond-1}(\xi_0)\Gamma_N^{\diamond}(\xi_0)\Sigma_N^{\diamond-1}(\xi_0) \xrightarrow{p} 0$. ■

Extensions of Unbalanced SPD Models and AQS Method

As discussed in the introduction and conclusion sections, the unbalanced SPD (USPD) models and the associated AQS methods are quite general in that they can be extended to allow for additional features in the model or to different types of unbalanced SPD models. For illustration, we extend the current model to allow errors to be serially correlated, and consider models with random effects (RE). We present some details on the following four extensions: (i) USPD model with two-way FE and serial correlation, (ii) USPD model with two-way FE, heteroskedasticity and serial correlation, (iii) USPD model with two-way RE and serial correlation, and (iv) USPD model with two-way RE, heteroskedasticity and serial correlation. For serial correlation, we assume that the model errors follow a stationary AR(1), i.e., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$. Cases (i), (ii) and (iv) all encounter incidental parameters problem, the standard methods for balanced panels cannot be applied, and the proposed AQS method needs to be called for. Case (iii) illustrates the simplicity of the proposed modeling strategy in controlling the random effects in the unbalanced SPD models with general time-

[11] The detail is tedeous and thus are omitted. Under balanced panel data model considered in footnote 7, we can easily show $\mathtt{diag}(\mathbb{L}_a\mathbb{P}_{\mathbb{D}}) = O(\frac{1}{n})$ for $a = \lambda, \rho$.

varying spatial weight matrices and serial correlation.

(i) USPD Model with Two-Way FE and Serial Correlation

Assume $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$, and $e_{it} \sim \text{iid}(0, \sigma_e^2)$. Denote $\mathbb{K} = \texttt{blkdiag}(D_1, \ldots, D_T)$. It is easy to see that $\text{Var}(\mathbf{V}) = \sigma_e^2 \mathbb{K}(\Omega_V(\varrho) \otimes I_n)\mathbb{K}' \equiv \sigma_e^2 \Upsilon_N(\varrho)$, where

$$\Omega_V(\varrho) = \frac{1}{1-\varrho^2} \begin{bmatrix} 1 & \varrho & \cdots & \varrho^{T-1} \\ \varrho & 1 & \cdots & \varrho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho^{T-1} & \varrho^{T-2} & & 1 \end{bmatrix}.$$

Denote $\mathbb{U}_N(\rho, \varrho) = \Upsilon_N^{-1}(\varrho) - \Upsilon_N^{-1}(\varrho)\mathbb{D}(\rho)[\mathbb{D}'(\rho)\Upsilon_N^{-1}(\varrho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\Upsilon_N^{-1}(\varrho)$. Let $\theta = (\beta', \sigma_e^2, \delta)'$, where $\delta = (\lambda, \rho, \varrho)'$. The concentrated quasi Gaussian loglikelihood function (with ϕ being concentrated) of θ takes the form:

$$\ell_N^c(\theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{N}{2}\ln \sigma_e^2 - \tfrac{1}{2}\ln|\Upsilon_N(\varrho)| + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_e^2}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta),$$

where $\tilde{\mathbf{V}}(\beta, \delta) = \mathbb{U}_N(\rho, \varrho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$. Hence, the concentrated quasi score (CQS) functions $S_N^c(\theta) = \frac{\partial}{\partial\theta}\ell_N^c(\theta)$ is given as

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_e^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N\sigma_e^2], \\ \frac{1}{\sigma_e^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{F}_N(\lambda)], \\ \frac{1}{\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{G}_N(\rho)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \frac{1}{2}\texttt{tr}[\Upsilon_N^{-1}(\varrho)\dot{\Upsilon}_N(\varrho)], \end{cases}$$

where $\dot{\Upsilon}_N(\varrho) = \frac{\partial}{\partial\varrho}\Upsilon_N(\varrho)$. To remove the effect from estimating FEs, we correct $S_N^c(\theta)$ using $S_N^*(\theta_0) = S_N^c(\theta_0) - \text{E}[S_N^c(\theta_0)]$, which takes the form at the general θ:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_e^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - \sigma_e^2\texttt{tr}(\mathbb{U}_N(\rho,\varrho))], \\ \frac{1}{\sigma_e^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho,\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)\Upsilon_N(\varrho)\mathbb{U}_N^2(\rho,\varrho)], \\ \frac{1}{\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\mathbb{U}_N(\rho,\varrho)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbf{V}}(\beta,\delta)'\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\beta,\delta) - \frac{1}{2}\texttt{tr}[\dot{\Upsilon}_N(\varrho)\mathbb{U}_N(\rho,\varrho)]. \end{cases}$$

Solving the AQS equations: $S_N^*(\theta) = 0$, gives the AQS estimator of θ.

(ii) USPD Model with Two-Way FE, Heteroskedasticity and Serial Correlation

Now, we consider the case that errors are heteroskedastic across individuals and serially correlated across time, i.e., $v_{it} = \varrho v_{i,t-1} + e_{it}$ with $|\varrho| < 1$ and $e_{it} \sim \text{inid}(0, \sigma_i^2)$. Let $h = \texttt{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. In this case, we have $\text{Var}(\mathbf{V}) = \mathbf{H}\Upsilon_N(\varrho)$, where $\mathbf{H} = \texttt{blkdiag}(h_1, \ldots, h_T)$ and h_t is obtained from h by omitting the rows and columns corresponding to the missing units at time t. Following the similar derivations as we do in Section 3.1, we obtain the desired AQS functions robust against the unknown heteroskedasticity:

$$S_N^{\diamond}(\beta, \delta) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{U}_N'(\rho, \varrho)\tilde{\mathbf{V}}(\beta, \delta), \\ \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\lambda, \rho) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho, \varrho) - \bar{\mathbb{G}}_N(\rho, \varrho)]\tilde{\mathbf{V}}(\beta, \delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{U}}_N(\rho, \varrho) - \bar{\mathbb{U}}_N(\rho, \varrho)]\tilde{\mathbf{V}}(\beta, \delta), \end{cases}$$

where $\bar{\mathbf{F}}_N'(\lambda, \rho) = \mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}_N'(\lambda)\mathbf{B}_N'(\rho)$, $\bar{\mathbb{F}}_N'(\delta) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{F}}_N'(\delta)\mathbb{U}_N(\rho, \varrho)]\texttt{diag}[\mathbb{U}_N(\rho, \varrho)]^{-1}$, $\bar{\mathbf{G}}_N(\rho, \varrho) = \mathbb{U}_N(\rho, \varrho)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)$, $\bar{\mathbb{G}}_N(\rho, \varrho) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{G}}_N(\rho, \varrho)\mathbb{U}_N(\rho, \varrho)]\texttt{diag}[\mathbb{U}_N(\rho, \varrho)]^{-1}$, $\bar{\mathbf{U}}_N(\rho, \varrho) = \mathbb{U}_N(\rho, \varrho)\dot{\Upsilon}_N(\varrho)$, and $\bar{\mathbb{U}}_N(\rho, \varrho) = \Upsilon_N^{-1}(\varrho)\texttt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{U}}_N(\rho, \varrho)\mathbb{U}_N(\rho, \varrho)]\texttt{diag}[\mathbb{U}_N(\rho, \varrho)]^{-1}$.

Solving the robust AQS equations: $S_N^{\diamond}(\beta, \delta) = 0$, gives the AQS estimators of β and δ, robust against unknown heteroskedasticity, and allowing serial correlation of AR(1) form.

(iii) USPD Model with Two-Way RE and Serial Correlation

Assume $\mu_i \sim \text{iid}(0, \sigma_\mu^2)$, $\alpha_t \sim \text{iid}(0, \sigma_\alpha^2)$, and they are mutually independent and independent of e_{it}. Then the covariance matrix of the composite error term is

$$\mho_N(\theta_1) = \sigma_\mu^2\mathcal{D}_\mu(\rho) + \sigma_\alpha^2\mathcal{D}_\alpha(\rho) + \sigma_e^2\Upsilon_N(\varrho),$$

with $\theta_1 = (\rho, \varrho, \sigma_e^2, \sigma_\mu^2, \sigma_\alpha^2)'$, $\mathcal{D}_\mu(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\mu\mathbf{D}_\mu'\mathbf{B}_N'(\rho)$ and $\mathcal{D}_\alpha(\rho) = \mathbf{B}_N(\rho)\mathbf{D}_\alpha\mathbf{D}_\alpha'\mathbf{B}_N'(\rho)$. Thus, the quasi Gaussian loglikelihood function of $\theta = (\beta', \lambda, \theta_1')'$ is

$$\ell_N(\theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{1}{2}\ln|\mho_N(\theta_1)| + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2}\mathbf{V}'(\beta, \lambda, \rho)\mho_N^{-1}(\theta_1)\mathbf{V}(\beta, \lambda, \rho),$$

where $\mathbf{V}(\beta, \lambda, \rho) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$. The direct QML estimator $\hat{\theta}_{\texttt{QML}}$ of θ maximizes the above equation $\ell_N(\theta)$, and its consistency and asymptotic normality can be easily established.

(iv) USPD Model with Two-Way RE, Heteroskedasticity and Serial Correlation

We now extend the model in (*iii*) to allow heteroskedasticity in the errors as in (*ii*) above.

Denote $\tilde{\mathbf{V}}(\theta) = \mho_N^{-1}(\theta_1)\mathbf{V}(\beta,\lambda,\rho)$. The quasi score functions assuming homoskedasticity are:

$$S_N(\theta) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\theta), \\ \mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\theta) - \mathtt{tr}[\mathbf{F}_N(\lambda)], \\ \sigma_e^2\tilde{\mathbf{V}}'(\theta)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\theta) - \sigma_e^2\mathtt{tr}[\mathbf{G}_N(\rho)\Upsilon_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{\sigma_e^2}{2}\tilde{\mathbf{V}}'(\theta)\dot{\Upsilon}_N(\varrho)\tilde{\mathbf{V}}(\theta) - \frac{\sigma_e^2}{2}\mathtt{tr}[\dot{\Upsilon}_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\Upsilon_N(\varrho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\Upsilon_N(\varrho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\mathcal{D}_\mu(\rho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\mathcal{D}_\mu(\rho)\mho_N^{-1}(\theta_1)], \\ \frac{1}{2}\tilde{\mathbf{V}}'(\theta)\mathcal{D}_\alpha(\rho)\tilde{\mathbf{V}}(\theta) - \frac{1}{2}\mathtt{tr}[\mathcal{D}_\alpha(\rho)\mho_N^{-1}(\theta_1)]. \end{cases}$$

It is easy to see that $\mathrm{E}[S_N^c(\theta_0)] \neq 0$ when e_{it}'s are heteroskedastic. Therefore, some adjustments on the above quasi score functions are necessary in order to have consistent estimation. Denote $\theta_2 = (\rho,\varrho,\sigma_\mu^2,\sigma_\alpha^2)'$, $\xi = (\beta',\lambda,\theta_2')'$, $\mho_N(\theta_2) = \sigma_\mu^2\mathcal{D}_\mu(\rho) + \sigma_\alpha^2\mathcal{D}_\alpha(\rho) + \Upsilon_N(\varrho)$ and $\tilde{\mathbf{V}}(\xi) = \mho_N^{-1}(\theta_2)\mathbf{V}(\beta,\lambda,\rho)$. Alone the similar ideas of Section, some tedious algebra leads to the AQS functions robust against the unknown heteroskedasticity and allowing serial correlation of AR(1) form:

$$S_N^\diamond(\xi) = \begin{cases} \mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\xi), \\ \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\lambda,\rho) - \bar{\mathbb{F}}_N'(\lambda,\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{F}}_N(\lambda,\rho) - \bar{\mathbb{F}}_N(\lambda,\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{G}}_N(\theta_2) - \bar{\mathbb{G}}_N(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{G}}_N(\theta_2) - \bar{\mathbb{G}}_N(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{U}}_N(\theta_2) - \bar{\mathbb{U}}_N(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{U}}_N(\theta_2) - \bar{\mathbb{U}}_N(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{S}}_\mu(\theta_2) - \bar{\mathbb{S}}_\mu(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{S}}_\mu(\theta_2) - \bar{\mathbb{S}}_\mu(\theta_2)], \\ \mathbf{V}'(\beta,\lambda,\rho)[\bar{\mathbf{S}}_\alpha(\theta_2) - \bar{\mathbb{S}}_\alpha(\theta_2)]\tilde{\mathbf{V}}(\xi) - \mathtt{tr}[\bar{\mathbf{S}}_\alpha(\theta_2) - \bar{\mathbb{S}}_\alpha(\theta_2)], \end{cases}$$

where $\bar{\mathbf{F}}_N'(\lambda,\rho) = \mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}_N'(\lambda)\mathbf{B}_N'(\rho)$, $\bar{\mathbb{F}}_N'(\lambda,\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{F}}_N'(\delta)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,

$\bar{\mathbf{G}}_N(\theta_2) = \mho_N^{-1}(\theta_2)\mathbf{G}_N(\rho)\Upsilon_N(\varrho)$, $\bar{\mathbb{G}}_N(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{G}}_N(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,

$\bar{\mathbf{U}}_N(\theta_2) = \mho_N^{-1}(\theta_2)\dot{\Upsilon}_N(\varrho)$, $\bar{\mathbb{U}}_N(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{U}}_N(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,

$\bar{\mathbf{S}}_\varpi(\theta_2) = \mho_N^{-1}(\theta_2)\mathcal{D}_\varpi(\rho)$, and $\bar{\mathbb{S}}_\varpi(\theta_2) = \Upsilon_N^{-1}(\varrho)\mathtt{diag}[\Upsilon_N(\varrho)\bar{\mathbf{S}}_\varpi(\theta_2)\mho_N^{-1}(\theta_2)]\mathtt{diag}[\mho_N^{-1}(\theta_2)]^{-1}$,

for $\varpi = \mu$ or α.

Solving the robust AQS equations: $S_N^\diamond(\xi) = 0$, gives the AQS estimator of ξ, robust against unknown heteroskedasticity, and allowing serial correlation of AR(1) form.

Asymptotic properties of the AQS estimators in cases (i) and (ii) can be studied in a similar way as that in the main text of the paper, and inferences methods can be developed along the same line. However, formal studies on these cases are still quite involved, and can only be done in a future research work. For the cases (iii) and (iv), we do not foresee any difficulties in establishing the asymptotic properties of the QML and AQS estimators, but developments of the inference methods may encounter some difficulties due to the involvement of three error components which may be all non-normal, and the allowance of unknown heteroskedasticity. Formal studies on these cases are in our future research agenda.

References

[1] Davidson, J., 1994. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

[2] Feller, W., 1968. *An introduction to probability theory and its applications*. New York: Wiley.