Supplementary Material to "Statistical Inference for Heavy-tailed Vector ARMA-GARCH/IGARCH Models"

Bibi Cai, Mengya Liu, Shiqing Ling*

Changsha University of Science and Technology,

Central China Normal University,

and Hong Kong University of Science and Technology

1 More simulation results

In this section, we conduct simulation experiments in a bivariate VARMA(1,1)-GARCH(1,1) model. We generate 1000 replications of sample size $n = 1000, 5000, 10000$ by the VARMA(1,1) - GARCH(1,1) model:

$$Y_t - \mu = \Phi(Y_{t-1} - \mu) + \varepsilon_t + \Theta\varepsilon_{t-1}, \tag{S.1}$$

$$\varepsilon_t = D_t\eta_t,\ H_t = W + A\vec{\varepsilon}_{t-1} + BH_{t-1}, \tag{S.2}$$

with a full matrix A and a diagonal matrix B. The parameter matrices in the ARMA part are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \Phi = \begin{pmatrix} 0.2 & 0.05 \\ 0.05 & 0.2 \end{pmatrix} \text{ and } \Theta = \begin{pmatrix} 0.7 & 0.05 \\ 0.05 & 0.7 \end{pmatrix}, \tag{S.3}$$

while the parameter matrices in the GARCH part are the same as (6.4)-(6.6) for three cases. The innovation η_t is a sequence of i.i.d. multivariate Gaussian distribution and standardized t_5 distribution, respectively, with mean 0 and covariance matrix Γ.

Tables S1-S6 list the sample biases (Bias), the asymptotic standard deviations (AD), and the sample standard deviations (SD) of $\hat{\lambda}_{sn}$ (SQMLE) and $\hat{\lambda}_n$ (LQMLE). Considering that the asymptotic normality of the local QMLE requires finite variance and IGARCH case, we hereafter exclude all simulation results concerning the LQMLE (including its AD and SD) for case where $E\left\|\varepsilon_t\right\|^{2\iota} = \infty$ with $\iota < 1$. From tables S1-S6, we find that the results are similar to the VAR(1)-GARCH(1,1) model: (i) in all cases, the AD and SD of both estimators are close to each other, which indicates our methods are reliable; (ii) the local QMLE has smaller AD and SD than those of self-weighted QMLE; (iii) Bias, AD, and SD for both estimators become smaller as sample size n increases.

We also assess the finite sample performance of the Wald tests W_{sn}

Table S1: Bias and standard deviations of estimators when $E\left\|\varepsilon_t\right\|^2 < \infty$ and $\eta_t \sim \mathcal{N}(0,\Gamma)$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0005	0.0375	0.0384	0.0000	0.0168	0.0165	-0.0002	0.0118	0.0118
	LQMLE	-0.0007	0.0373	0.0380	-0.0001	0.0167	0.0164	-0.0003	0.0118	0.0117
μ_2	SQMLE	-0.0011	0.0375	0.0365	-0.0006	0.0168	0.0164	-0.0007	0.0118	0.0114
	LQMLE	-0.0014	0.0372	0.0362	-0.0007	0.0167	0.0164	-0.0008	0.0118	0.0113
Φ_{11}	SQMLE	-0.0002	0.0434	0.0445	0.0005	0.0196	0.0194	0.0004	0.0138	0.0144
	LQMLE	-0.0008	0.0418	0.0427	0.0005	0.0189	0.0188	0.0004	0.0133	0.0138
Φ_{21}	SQMLE	0.0023	0.0432	0.0433	0.0008	0.0194	0.0192	0.0000	0.0138	0.0139
	LQMLE	0.0017	0.0417	0.0411	0.0004	0.0187	0.0188	-0.0001	0.0133	0.0134
Φ_{12}	SQMLE	0.0018	0.0432	0.0429	0.0008	0.0194	0.0196	-0.0008	0.0137	0.0135
	LQMLE	0.0019	0.0416	0.0407	0.0007	0.0187	0.0191	-0.0007	0.0132	0.0130
Φ_{22}	SQMLE	-0.0018	0.0435	0.0435	0.0005	0.0196	0.0200	-0.0003	0.0139	0.0139
	LQMLE	-0.0021	0.0419	0.0415	0.0006	0.0189	0.0191	-0.0001	0.0133	0.0136
Θ_{11}	SQMLE	0.0002	0.0293	0.0300	0.0003	0.0132	0.0129	0.0000	0.0093	0.0095
	LQMLE	0.0004	0.0287	0.0293	0.0000	0.0130	0.0127	0.0000	0.0092	0.0093
Θ_{21}	SQMLE	-0.0033	0.0292	0.0290	-0.0008	0.0131	0.0131	0.0002	0.0093	0.0095
	LQMLE	-0.0032	0.0287	0.0287	-0.0006	0.0129	0.0129	0.0002	0.0091	0.0094
Θ_{12}	SQMLE	0.0001	0.0292	0.0293	-0.0010	0.0132	0.0128	0.0001	0.0093	0.0095
	LQMLE	0.0000	0.0287	0.0288	-0.0009	0.0129	0.0127	0.0001	0.0091	0.0093
Θ_{22}	SQMLE	0.0010	0.0293	0.0291	-0.0002	0.0132	0.0132	0.0004	0.0093	0.0097
	LQMLE	0.0010	0.0288	0.0287	-0.0002	0.0130	0.0131	0.0003	0.0092	0.0094
W_1	SQMLE	0.0017	0.0133	0.0141	0.0000	0.0058	0.0058	0.0001	0.0041	0.0040
	LQMLE	0.0015	0.0131	0.0139	-0.0001	0.0057	0.0057	0.0001	0.0041	0.0039
W_2	SQMLE	0.0023	0.0134	0.0142	0.0003	0.0058	0.0060	0.0002	0.0041	0.0041
	LQMLE	0.0021	0.0131	0.0140	0.0003	0.0057	0.0059	0.0001	0.0041	0.0041
A_{11}	SQMLE	-0.0024	0.0426	0.0442	-0.0009	0.0193	0.0191	-0.0007	0.0137	0.0135
	LQMLE	-0.0021	0.0408	0.0429	-0.0010	0.0185	0.0183	-0.0006	0.0131	0.0130
A_{21}	SQMLE	0.0045	0.0412	0.0442	-0.0007	0.0184	0.0183	0.0002	0.0131	0.0133
	LQMLE	0.0046	0.0395	0.0417	-0.0004	0.0177	0.0176	0.0003	0.0126	0.0129
A_{12}	SQMLE	0.0010	0.0407	0.0437	0.0006	0.0185	0.0180	0.0006	0.0131	0.0131
	LQMLE	0.0013	0.0390	0.0416	0.0007	0.0178	0.0175	0.0006	0.0125	0.0126
A_{22}	SQMLE	-0.0023	0.0429	0.0459	-0.0021	0.0193	0.0190	-0.0001	0.0137	0.0136
	LQMLE	-0.0026	0.0411	0.0437	-0.0018	0.0185	0.0183	0.0000	0.0131	0.0132
B_{11}	SQMLE	-0.0037	0.0576	0.0606	0.0003	0.0259	0.0253	-0.0005	0.0183	0.0180
	LQMLE	-0.0035	0.0562	0.0608	0.0004	0.0253	0.0249	-0.0005	0.0179	0.0177
B_{22}	SQMLE	-0.0072	0.0581	0.0618	0.0007	0.0259	0.0266	-0.0003	0.0183	0.0185
	LQMLE	-0.0068	0.0568	0.0601	0.0005	0.0253	0.0260	-0.0003	0.0179	0.0183
σ_{21}	SQMLE	0.0010	0.0293	0.0294	0.0001	0.0131	0.0132	0.0002	0.0093	0.0096
	LQMLE	0.0011	0.0287	0.0287	0.0001	0.0129	0.0129	0.0001	0.0091	0.0094

and W_n as in (3.7) and (4.2) of VARMA(1,1)-GARCH(1,1) model. For the linear Wald test for the ARMA part, the setting of ARMA parameter matrices are set to be

$$\mu = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \Phi = \begin{pmatrix} 0.2 & \kappa \\ 0.05 & 0.2 \end{pmatrix} \text{ and } \Theta = \begin{pmatrix} 0.7 & 0.05 \\ 0.05 & 0.7 \end{pmatrix},$$

and the GARCH parameter matrices are the same as (6.4)-(6.6). $\kappa = 0, 0.05, 0.1$ in all cases, where the null hypothesis is $\kappa = 0$. We fit each

Table S2: Bias and standard deviations of estimators when $E\|\varepsilon_t\|^2 < \infty$ and $\eta_t \sim t_5$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0002	0.0368	0.0380	-0.0001	0.0173	0.0178	-0.0004	0.0123	0.0121
	LQMLE	-0.0007	0.0368	0.0384	-0.0002	0.0173	0.0179	-0.0003	0.0123	0.0121
μ_2	SQMLE	-0.0008	0.0368	0.0380	0.0004	0.0172	0.0174	0.0000	0.0123	0.0121
	LQMLE	-0.0012	0.0368	0.0382	0.0004	0.0172	0.0174	0.0000	0.0123	0.0121
Φ_{11}	SQMLE	-0.0012	0.0520	0.0540	-0.0005	0.0248	0.0257	-0.0002	0.0177	0.0175
	LQMLE	-0.0024	0.0498	0.0515	-0.0007	0.0237	0.0245	-0.0003	0.0169	0.0167
Φ_{21}	SQMLE	-0.0013	0.0520	0.0572	-0.0004	0.0244	0.0252	-0.0005	0.0175	0.0175
	LQMLE	-0.0017	0.0498	0.0543	-0.0004	0.0233	0.0241	-0.0003	0.0168	0.0169
Φ_{12}	SQMLE	-0.0001	0.0513	0.0521	0.0008	0.0243	0.0254	0.0003	0.0174	0.0176
	LQMLE	-0.0001	0.0490	0.0511	0.0008	0.0233	0.0248	0.0004	0.0166	0.0167
Φ_{22}	SQMLE	-0.0013	0.0522	0.0549	0.0003	0.0245	0.0253	-0.0001	0.0176	0.0178
	LQMLE	-0.0021	0.0500	0.0516	0.0002	0.0234	0.0242	0.0001	0.0168	0.0171
Θ_{11}	SQMLE	0.0007	0.0340	0.0368	0.0000	0.0162	0.0171	0.0002	0.0115	0.0113
	LQMLE	0.0012	0.0331	0.0360	0.0001	0.0158	0.0166	0.0002	0.0112	0.0109
Θ_{21}	SQMLE	0.0002	0.0339	0.0378	-0.0003	0.0161	0.0167	0.0002	0.0115	0.0111
	LQMLE	0.0002	0.0331	0.0365	-0.0002	0.0156	0.0163	0.0001	0.0112	0.0109
Θ_{12}	SQMLE	-0.0008	0.0336	0.0349	-0.0003	0.0159	0.0161	-0.0001	0.0114	0.0108
	LQMLE	-0.0010	0.0328	0.0342	-0.0005	0.0156	0.0156	-0.0001	0.0111	0.0106
Θ_{22}	SQMLE	-0.0001	0.0340	0.0369	-0.0004	0.0161	0.0164	-0.0004	0.0115	0.0112
	LQMLE	0.0004	0.0332	0.0358	-0.0005	0.0157	0.0158	-0.0004	0.0112	0.0109
W_1	SQMLE	0.0022	0.0162	0.0182	0.0011	0.0080	0.0088	0.0000	0.0058	0.0060
	LQMLE	0.0018	0.0158	0.0176	0.0009	0.0078	0.0086	-0.0001	0.0056	0.0058
W_2	SQMLE	0.0025	0.0162	0.0192	0.0007	0.0080	0.0086	0.0006	0.0059	0.0064
	LQMLE	0.0020	0.0159	0.0190	0.0006	0.0078	0.0084	0.0005	0.0058	0.0063
A_{11}	SQMLE	0.0016	0.0670	0.0785	0.0041	0.0354	0.0466	0.0006	0.0249	0.0267
	LQMLE	0.0020	0.0630	0.0730	0.0034	0.0330	0.0425	0.0005	0.0232	0.0250
A_{21}	SQMLE	0.0121	0.0669	0.0839	0.0030	0.0326	0.0357	0.0027	0.0236	0.0265
	LQMLE	0.0120	0.0624	0.0774	0.0033	0.0306	0.0337	0.0025	0.0221	0.0246
A_{12}	SQMLE	0.0066	0.0633	0.0743	0.0021	0.0321	0.0342	0.0022	0.0236	0.0272
	LQMLE	0.0060	0.0590	0.0698	0.0025	0.0301	0.0318	0.0017	0.0222	0.0254
A_{22}	SQMLE	0.0077	0.0703	0.0918	-0.0008	0.0334	0.0367	0.0004	0.0243	0.0255
	LQMLE	0.0075	0.0659	0.0848	-0.0010	0.0313	0.0341	0.0005	0.0228	0.0239
B_{11}	SQMLE	-0.0147	0.0794	0.0932	-0.0070	0.0403	0.0449	-0.0014	0.0297	0.0320
	LQMLE	-0.0140	0.0767	0.0905	-0.0064	0.0389	0.0431	-0.0011	0.0286	0.0307
B_{22}	SQMLE	-0.0209	0.0802	0.0966	-0.0037	0.0405	0.0435	-0.0039	0.0296	0.0309
	LQMLE	-0.0208	0.0775	0.0967	-0.0036	0.0391	0.0422	-0.0037	0.0286	0.0305
σ_{21}	SQMLE	0.0007	0.0315	0.0327	0.0006	0.0145	0.0150	-0.0001	0.0103	0.0105
	LQMLE	0.0007	0.0308	0.0322	0.0007	0.0141	0.0146	-0.0002	0.0101	0.0102

replication by a VARMA(1,1)-GARCH(1,1) model, and then use W_{sn} and W_n to detect the null hypothesis. The nonlinear VIGARCH model tests (4.4) and (4.5) are conducted for the VARMA(1,1)-GARCH(1,1) models (S.1) and (S.2) with the parameter matrices as (S.3) and (6.4)-(6.6). Tables S7-S8 reports the percentages of the empirical size and the power of the Wald tests, respectively.

From Tables S7-S8, (i) for the ARMA part, the sizes of W_{sn} and W_n

Table S3: Bias and standard deviations of estimators for VIGARCH case and $\eta_t \sim \mathcal{N}(0,\Gamma)$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0007	0.0459	0.0474	0.0000	0.0204	0.0200	-0.0002	0.0144	0.0144
	LQMLE	-0.0008	0.0458	0.0471	0.0000	0.0203	0.0200	-0.0003	0.0144	0.0143
μ_2	SQMLE	-0.0013	0.0458	0.0446	-0.0007	0.0204	0.0199	-0.0009	0.0144	0.0139
	LQMLE	-0.0014	0.0457	0.0444	-0.0008	0.0203	0.0199	-0.0009	0.0144	0.0139
Φ_{11}	SQMLE	-0.0007	0.0431	0.0444	0.0004	0.0194	0.0193	0.0005	0.0137	0.0144
	LQMLE	-0.0011	0.0413	0.0423	0.0006	0.0186	0.0187	0.0004	0.0132	0.0136
Φ_{21}	SQMLE	0.0013	0.0421	0.0428	0.0007	0.0189	0.0190	0.0000	0.0134	0.0135
	LQMLE	0.0008	0.0404	0.0405	0.0003	0.0181	0.0183	-0.0001	0.0128	0.0130
Φ_{12}	SQMLE	0.0014	0.0421	0.0419	0.0007	0.0189	0.0190	-0.0007	0.0134	0.0133
	LQMLE	0.0017	0.0403	0.0397	0.0007	0.0181	0.0186	-0.0007	0.0128	0.0126
Φ_{22}	SQMLE	-0.0020	0.0432	0.0428	0.0005	0.0194	0.0197	-0.0003	0.0137	0.0137
	LQMLE	-0.0023	0.0414	0.0410	0.0006	0.0186	0.0188	-0.0002	0.0132	0.0134
Θ_{11}	SQMLE	0.0003	0.0288	0.0297	0.0003	0.0130	0.0127	-0.0001	0.0092	0.0095
	LQMLE	0.0005	0.0281	0.0288	0.0001	0.0127	0.0125	-0.0001	0.0090	0.0091
Θ_{21}	SQMLE	-0.0025	0.0284	0.0284	-0.0008	0.0128	0.0128	0.0001	0.0091	0.0092
	LQMLE	-0.0023	0.0277	0.0280	-0.0006	0.0125	0.0125	0.0002	0.0088	0.0090
Θ_{12}	SQMLE	0.0001	0.0284	0.0283	-0.0011	0.0128	0.0124	0.0001	0.0090	0.0092
	LQMLE	0.0000	0.0277	0.0277	-0.0010	0.0125	0.0122	0.0001	0.0088	0.0090
Θ_{22}	SQMLE	0.0010	0.0289	0.0290	-0.0001	0.0130	0.0130	0.0004	0.0092	0.0095
	LQMLE	0.0011	0.0282	0.0285	-0.0001	0.0127	0.0128	0.0003	0.0090	0.0093
W_1	SQMLE	0.0025	0.0148	0.0153	-0.0001	0.0064	0.0061	0.0003	0.0045	0.0044
	LQMLE	0.0023	0.0145	0.0153	-0.0001	0.0062	0.0060	0.0002	0.0044	0.0043
W_2	SQMLE	0.0028	0.0147	0.0157	0.0005	0.0064	0.0067	0.0002	0.0045	0.0044
	LQMLE	0.0026	0.0144	0.0155	0.0004	0.0063	0.0066	0.0002	0.0044	0.0044
A_{11}	SQMLE	-0.0012	0.0464	0.0483	-0.0008	0.0210	0.0207	-0.0006	0.0149	0.0146
	LQMLE	-0.0019	0.0443	0.0461	-0.0010	0.0200	0.0197	-0.0006	0.0142	0.0140
A_{21}	SQMLE	0.0052	0.0392	0.0418	-0.0005	0.0175	0.0173	0.0000	0.0124	0.0126
	LQMLE	0.0047	0.0373	0.0396	-0.0004	0.0167	0.0164	0.0002	0.0118	0.0121
A_{12}	SQMLE	0.0017	0.0388	0.0414	0.0006	0.0176	0.0170	0.0004	0.0124	0.0124
	LQMLE	0.0016	0.0370	0.0395	0.0005	0.0167	0.0163	0.0005	0.0118	0.0118
A_{22}	SQMLE	-0.0018	0.0467	0.0490	-0.0020	0.0210	0.0207	0.0001	0.0149	0.0147
	LQMLE	-0.0026	0.0445	0.0469	-0.0020	0.0200	0.0197	0.0001	0.0142	0.0141
B_{11}	SQMLE	-0.0038	0.0472	0.0494	0.0003	0.0212	0.0201	-0.0006	0.0150	0.0147
	LQMLE	-0.0034	0.0460	0.0492	0.0005	0.0206	0.0196	-0.0005	0.0145	0.0142
B_{22}	SQMLE	-0.0060	0.0473	0.0500	0.0006	0.0212	0.0215	-0.0002	0.0150	0.0149
	LQMLE	-0.0055	0.0461	0.0490	0.0007	0.0206	0.0209	-0.0002	0.0145	0.0147
σ_{21}	SQMLE	0.0013	0.0294	0.0295	0.0003	0.0132	0.0133	0.0002	0.0094	0.0097
	LQMLE	0.0012	0.0287	0.0288	0.0001	0.0129	0.0129	0.0001	0.0091	0.0094

approach the nominal ones as n increases, while the size of W_{sn} are a little less than the nominal ones when the sample size n=10000 and the case $E\|\varepsilon_t\|^{2\iota} = \infty$; (ii) for testing VIGARCH model, the sizes of W_{sn} are close to the nominal ones as n increases, while the sizes of W_n are little larger than the nominal ones; (iv) for all tests, the powers of all the tests increase as the value of n increases; (v) for all tests, the powers of Wald tests when $\eta_t \sim \mathcal{N}(0,\Gamma)$ are larger than the powers of Wald tests when $\eta_t \sim t_5$.

Table S4: Bias and standard deviations of estimators for VIGARCH case and $\eta_t \sim t_5$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0002	0.0423	0.0438	0.0001	0.0200	0.0207	-0.0005	0.0142	0.0140
	LQMLE	-0.0005	0.0423	0.0441	0.0000	0.0200	0.0207	-0.0004	0.0142	0.0140
μ_2	SQMLE	-0.0007	0.0423	0.0442	0.0006	0.0199	0.0203	-0.0001	0.0142	0.0141
	LQMLE	-0.0009	0.0424	0.0443	0.0005	0.0199	0.0203	0.0000	0.0143	0.0141
Φ_{11}	SQMLE	-0.0015	0.0530	0.0551	-0.0010	0.0255	0.0261	-0.0002	0.0182	0.0180
	LQMLE	-0.0025	0.0507	0.0523	-0.0009	0.0243	0.0248	-0.0001	0.0174	0.0172
Φ_{21}	SQMLE	-0.0011	0.0517	0.0568	-0.0003	0.0245	0.0250	-0.0004	0.0176	0.0176
	LQMLE	-0.0012	0.0493	0.0537	-0.0003	0.0233	0.0238	-0.0003	0.0168	0.0170
Φ_{12}	SQMLE	-0.0003	0.0510	0.0518	0.0009	0.0244	0.0251	0.0002	0.0175	0.0176
	LQMLE	-0.0002	0.0486	0.0504	0.0007	0.0231	0.0252	0.0004	0.0166	0.0166
Φ_{22}	SQMLE	-0.0019	0.0535	0.0559	0.0002	0.0253	0.0261	-0.0002	0.0182	0.0185
	LQMLE	-0.0024	0.0511	0.0527	0.0000	0.0241	0.0248	0.0000	0.0173	0.0176
Θ_{11}	SQMLE	0.0007	0.0343	0.0374	0.0002	0.0165	0.0169	0.0001	0.0118	0.0116
	LQMLE	0.0011	0.0334	0.0365	0.0003	0.0160	0.0165	0.0001	0.0114	0.0112
Θ_{21}	SQMLE	0.0003	0.0336	0.0373	-0.0002	0.0161	0.0167	0.0001	0.0115	0.0112
	LQMLE	0.0003	0.0327	0.0360	-0.0002	0.0156	0.0162	0.0001	0.0112	0.0109
Θ_{12}	SQMLE	-0.0008	0.0334	0.0353	-0.0003	0.0160	0.0158	0.0000	0.0114	0.0108
	LQMLE	-0.0009	0.0325	0.0343	-0.0004	0.0155	0.0157	-0.0001	0.0111	0.0107
Θ_{22}	SQMLE	-0.0003	0.0345	0.0374	-0.0003	0.0165	0.0168	-0.0004	0.0118	0.0115
	LQMLE	0.0002	0.0335	0.0363	-0.0003	0.0160	0.0161	-0.0004	0.0115	0.0111
W_1	SQMLE	0.0024	0.0168	0.0186	0.0010	0.0083	0.0090	0.0000	0.0060	0.0062
	LQMLE	0.0021	0.0165	0.0182	0.0008	0.0081	0.0088	0.0000	0.0058	0.0060
W_2	SQMLE	0.0023	0.0168	0.0196	0.0008	0.0083	0.0088	0.0006	0.0061	0.0066
	LQMLE	0.0019	0.0165	0.0193	0.0007	0.0081	0.0086	0.0005	0.0060	0.0065
A_{11}	SQMLE	0.0028	0.0753	0.0902	0.0040	0.0391	0.0445	0.0008	0.0276	0.0297
	LQMLE	0.0024	0.0711	0.0835	0.0035	0.0366	0.0456	0.0008	0.0260	0.0277
A_{21}	SQMLE	0.0098	0.0630	0.0756	0.0024	0.0310	0.0326	0.0022	0.0226	0.0258
	LQMLE	0.0092	0.0587	0.0704	0.0027	0.0291	0.0312	0.0020	0.0211	0.0238
A_{12}	SQMLE	0.0049	0.0599	0.0693	0.0020	0.0307	0.0327	0.0020	0.0225	0.0256
	LQMLE	0.0047	0.0561	0.0655	0.0023	0.0287	0.0307	0.0015	0.0211	0.0240
A_{22}	SQMLE	0.0085	0.0775	0.0923	-0.0002	0.0372	0.0411	0.0004	0.0270	0.0281
	LQMLE	0.0075	0.0729	0.0849	-0.0008	0.0349	0.0381	0.0004	0.0255	0.0266
B_{11}	SQMLE	-0.0121	0.0674	0.0780	-0.0057	0.0344	0.0373	-0.0015	0.0252	0.0273
	LQMLE	-0.0121	0.0653	0.0762	-0.0054	0.0331	0.0364	-0.0013	0.0243	0.0261
B_{22}	SQMLE	-0.0173	0.0682	0.0794	-0.0035	0.0341	0.0368	-0.0029	0.0251	0.0257
	LQMLE	-0.0169	0.0661	0.0783	-0.0032	0.0329	0.0353	-0.0027	0.0242	0.0253
σ_{21}	SQMLE	0.0007	0.0315	0.0329	0.0006	0.0145	0.0151	-0.0001	0.0103	0.0105
	LQMLE	0.0007	0.0308	0.0324	0.0007	0.0141	0.0146	-0.0002	0.0101	0.0102

To access the performance of portmanteau tests $Q_{sn}(M)$ and $Q_n(M)$ in the finite samples, we generate 1000 replications of sample size $n = 1000, 5000, 10000$ from two different models. The null model is model (S.1)-(S.2) with parameter sets (S.3) and (6.4)-(6.6), denoted by Model 1. The alternative models are as follows:

$$\textbf{Model 2}: \quad Y_t - \mu = \Phi(Y_{t-1} - \mu) + \varepsilon_t + \Theta\varepsilon_{t-1},$$

Table S5: Bias and standard deviations of estimators when $E\|\varepsilon_t\|^{2\iota} = \infty$ and $\eta_t \sim \mathcal{N}(0, \Gamma)$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0009	0.0712	0.0775	0.0000	0.0304	0.0303	-0.0001	0.0215	0.0215
μ_2	SQMLE	-0.0026	0.0711	0.0699	-0.0004	0.0304	0.0299	-0.0013	0.0215	0.0213
Φ_{11}	SQMLE	-0.0007	0.0431	0.0445	0.0002	0.0194	0.0195	0.0006	0.0138	0.0144
Φ_{21}	SQMLE	0.0011	0.0422	0.0423	0.0005	0.0189	0.0190	-0.0001	0.0134	0.0137
Φ_{12}	SQMLE	0.0012	0.0420	0.0414	0.0007	0.0189	0.0190	-0.0006	0.0134	0.0131
Φ_{22}	SQMLE	-0.0021	0.0432	0.0432	0.0004	0.0194	0.0194	-0.0002	0.0138	0.0139
Θ_{11}	SQMLE	0.0000	0.0291	0.0298	0.0004	0.0132	0.0130	-0.0001	0.0093	0.0095
Θ_{21}	SQMLE	-0.0026	0.0286	0.0289	-0.0006	0.0129	0.0129	0.0001	0.0091	0.0094
Θ_{12}	SQMLE	-0.0001	0.0286	0.0284	-0.0010	0.0129	0.0127	0.0001	0.0091	0.0092
Θ_{22}	SQMLE	0.0007	0.0291	0.0293	-0.0001	0.0132	0.0133	0.0003	0.0093	0.0096
W_1	SQMLE	0.0040	0.0216	0.0206	0.0003	0.0088	0.0084	0.0006	0.0062	0.0062
W_2	SQMLE	0.0046	0.0217	0.0214	0.0010	0.0088	0.0090	0.0004	0.0062	0.0062
A_{11}	SQMLE	-0.0005	0.0444	0.0459	-0.0003	0.0200	0.0199	-0.0003	0.0142	0.0140
A_{21}	SQMLE	0.0057	0.0373	0.0396	-0.0001	0.0166	0.0163	0.0001	0.0118	0.0119
A_{12}	SQMLE	0.0020	0.0368	0.0390	0.0010	0.0167	0.0163	0.0005	0.0118	0.0118
A_{22}	SQMLE	-0.0012	0.0445	0.0469	-0.0016	0.0200	0.0197	0.0004	0.0142	0.0142
B_{11}	SQMLE	-0.0021	0.0399	0.0412	-0.0003	0.0180	0.0176	-0.0006	0.0127	0.0125
B_{22}	SQMLE	-0.0038	0.0399	0.0415	0.0004	0.0179	0.0176	-0.0002	0.0127	0.0129
σ_{21}	SQMLE	0.0020	0.0296	0.0298	0.0004	0.0133	0.0136	0.0002	0.0094	0.0098

Table S6: Bias and standard deviations of estimators when $E\|\varepsilon_t\|^{2\iota} = \infty$ and $\eta_t \sim t_5$ for VARMA(1,1)-GARCH(1,1) model

parameter	Estimator	n=1000			n=5000			n=10000		
		Bias	AD	SD	Bias	AD	SD	Bias	AD	SD
μ_1	SQMLE	-0.0006	0.0537	0.0567	0.0001	0.0253	0.0264	-0.0007	0.0180	0.0178
μ_2	SQMLE	-0.0016	0.0537	0.0577	0.0008	0.0253	0.0255	-0.0002	0.0181	0.0179
Φ_{11}	SQMLE	-0.0013	0.0535	0.0552	-0.0007	0.0258	0.0269	-0.0003	0.0184	0.0182
Φ_{21}	SQMLE	-0.0013	0.0520	0.0567	-0.0002	0.0247	0.0254	-0.0003	0.0178	0.0179
Φ_{12}	SQMLE	-0.0009	0.0513	0.0522	0.0007	0.0246	0.0267	0.0001	0.0177	0.0176
Φ_{22}	SQMLE	-0.0014	0.0540	0.0561	0.0000	0.0256	0.0260	-0.0003	0.0183	0.0186
Θ_{11}	SQMLE	0.0004	0.0350	0.0383	0.0000	0.0169	0.0175	0.0001	0.0120	0.0117
Θ_{21}	SQMLE	0.0005	0.0341	0.0381	-0.0002	0.0164	0.0170	0.0001	0.0117	0.0113
Θ_{12}	SQMLE	-0.0005	0.0339	0.0357	-0.0003	0.0162	0.0166	0.0000	0.0116	0.0108
Θ_{22}	SQMLE	0.0000	0.0350	0.0382	-0.0002	0.0168	0.0171	-0.0003	0.0120	0.0116
W_1	SQMLE	0.0022	0.0195	0.0207	0.0007	0.0096	0.0105	0.0001	0.0069	0.0072
W_2	SQMLE	0.0023	0.0197	0.0215	0.0010	0.0096	0.0100	0.0006	0.0071	0.0075
A_{11}	SQMLE	0.0021	0.0722	0.0851	0.0033	0.0374	0.0454	0.0009	0.0264	0.0283
A_{21}	SQMLE	0.0090	0.0596	0.0726	0.0030	0.0297	0.0325	0.0020	0.0215	0.0248
A_{12}	SQMLE	0.0041	0.0565	0.0638	0.0020	0.0291	0.0312	0.0020	0.0214	0.0245
A_{22}	SQMLE	0.0087	0.0750	0.0934	0.0001	0.0355	0.0392	0.0003	0.0258	0.0268
B_{11}	SQMLE	-0.0081	0.0581	0.0641	-0.0040	0.0299	0.0322	-0.0016	0.0218	0.0235
B_{22}	SQMLE	-0.0137	0.0596	0.0697	-0.0033	0.0295	0.0322	-0.0022	0.0215	0.0220
σ_{21}	SQMLE	0.0008	0.0317	0.0327	0.0008	0.0146	0.0151	-0.0002	0.0104	0.0106

$$\varepsilon_t = D_t\eta_t,\ H_t = W + A_1\vec{\varepsilon}_{t-1} + A_2\vec{\varepsilon}_{t-2} + BH_{t-1}.$$

where the parameter matrices in the ARMA part are the same as (S.3), with the parameter matrices in the GARCH part are the same as (6.8)-(6.10).

Table S7: Size and power of the Wald tests for ARMA part for VARMA(1,1)-GARCH(1,1) model

		$E\|\varepsilon_t\|^2<\infty$						$VIGARCH$						$E\|\varepsilon_t\|^{2\iota}=\infty$		
		$\kappa=0.0$		$\kappa=0.05$		$\kappa=0.1$		$\kappa=0.0$		$\kappa=0.05$		$\kappa=0.1$		$\kappa=0.0$	$\kappa=0.05$	$\kappa=0.1$
η_t	n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_{sn}	W_{sn}
$\mathcal{N}(0,\Gamma)$	1000	4.8	5.3	22.3	22.5	63.9	66.8	6.1	5.5	23.3	24.2	65.5	69.3	4.7	23.4	64.6
	5000	5.3	5.4	74.6	77.6	100	99.9	5.3	5.7	77.5	80.7	100	100	5	76.9	100
	10000	4.5	4.2	94.9	96.6	100	100	4.8	4.4	95.8	97.6	100	100	3.9	96.5	100
t_5	1000	5.6	5.9	18.7	20	48.5	52.8	5.1	6.3	19	19.6	50.1	53.8	5.1	18.7	48.9
	5000	4.9	5.7	56.1	60	97.4	97.9	4.2	5.4	56.6	60.7	97.1	98	4.6	55.9	97.1
	10000	3.7	4.9	81.4	86.1	99.9	100	4.3	4.7	81.9	86.5	99.9	99.9	3.3	81.6	99.9

Table S8: Size and power of Wald tests for VIGARCH errors for VARMA(1,1)-GARCH(1,1) model

	Size with $VIGARCH$				Power with $E\|\varepsilon_t\|^2<\infty$				Power with $E\|\varepsilon_t\|^{2\iota}=\infty$	
	$N(0,\Gamma)$		t_5		$N(0,\Gamma)$		t_5		$N(0,\Gamma)$	t_5
n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_n	W_{sn}	W_{sn}
1000	5.1	6.5	3	5.3	72.7	76.2	29.6	34	54.1	12.2
5000	4.2	5.6	5.7	7.2	100	100	85.1	87.1	99.4	64.7
10000	5.5	6.9	4.3	6.7	100	100	96.2	97.5	100	93.3

All simulated data from Models 1 and 2 are fitted by VARMA(1,1)-GARCH(1,1) model. M is set to be 6 for all cases. The results are summarized in Table S9, and the sizes correspond to the Model 1. From the table, we can show that:

Table S9: Size and power of portmanteau tests

		$E\|\varepsilon_t\|^2<\infty$				$VIGARCH$				$E\|\varepsilon_t\|^{2\iota}=\infty$	
		$\mathcal{N}(0,\Gamma)$		t_5		$\mathcal{N}(0,\Gamma)$		t_5		$\mathcal{N}(0,\Gamma)$	t_5
	n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_n	Q_{sn}	Q_{sn}
Model 1	1000	7	7	9.8	6.1	6.1	6.2	8.3	5.6	5.8	7.6
	5000	5.6	5.1	9.1	5.2	5	5.5	7.9	5.6	5.5	7.4
	10000	5.8	5.9	7.4	6.2	5.7	5.9	6.4	6.2	5.8	6.4
Model 2	1000	75.7	81.5	32.1	29.4	72.4	77.7	29.2	27.3	85.6	34.1
	5000	100	100	86.7	77.9	100	100	82.5	74.6	100	90.5
	10000	100	100	98.1	94.4	100	100	98.1	92.3	100	99.1

From the Table S9, we can show that: (i) the sizes of Q_{sn} and Q_n are a little larger than their nominal ones and close to the nominal ones as sample size n increase, except for some cases; (ii) all the powers increase when n increases; (iii) the powers when $\eta \sim \mathcal{N}(0,\Gamma)$ are greater than the powers when $\eta_t \sim t_5$.

2 Technical Proofs

We give some basic formulas as follows:

$$\frac{\partial \varepsilon_t(\varphi)}{\partial \varphi'} = -\Psi^{-1}(L)[\Phi(1), X_{t-1}\otimes I_m], \tag{S.4}$$

$$\frac{\partial H_t(\lambda)}{\partial \delta'} = \Big(I_m - \sum_{i=1}^{s} B_i L^i\Big)^{-1}\Big(I_m, \tilde{H}_{t-1}\otimes I_m\Big), \tag{S.5}$$

$$\frac{\partial H_t(\lambda)}{\partial \varphi'} = \Big(I_m - \sum_{i=1}^{s} B_i L^i\Big)^{-1}\Big(\sum_{i=1}^{r} A_i L^i\Big)\Big(2\vec{\varepsilon}_t^{\,*}\frac{\partial \varepsilon_t(\varphi)}{\partial \varphi'}\Big), \tag{S.6}$$

where $X_{t-1} = (Y'_{t-1}-\mu', \ldots, Y'_{t-p}-\mu', \varepsilon'_{t-1}(\varphi), \ldots, \varepsilon'_{t-q}(\varphi))$, $\tilde{H}_{t-1} = (\vec{\varepsilon}'_{t-1}(\varphi), \ldots, \vec{\varepsilon}'_{t-r}(\varphi), H'_{t-1}(\lambda), \ldots, H'_{t-s}(\lambda))$, and $\vec{\varepsilon}_t^{\,*}(\varphi) = \mathrm{diag}(\varepsilon_{1t}(\varphi), \ldots, \varepsilon_{mt}(\varphi))$.

We further give the quasi-score function and the quasi-information matrix as follows.

$$\frac{\partial l_t(\lambda)}{\partial \varphi} = -\frac{\partial \varepsilon'_t(\varphi)}{\partial \varphi}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_t(\varphi) - \frac{1}{2}\frac{\partial H'_t(\lambda)}{\partial \varphi}D_t^{-2}(\lambda)\zeta_t(\lambda), \tag{S.7}$$

$$\frac{\partial l_t(\lambda)}{\partial \delta} = -\frac{1}{2}\frac{\partial H'_t(\lambda)}{\partial \delta}D_t^{-2}(\lambda)\zeta_t(\lambda), \tag{S.8}$$

$$\frac{\partial l_t(\lambda)}{\partial \sigma} = -\frac{1}{2}\frac{\partial \mathrm{vec}'(\Gamma)}{\partial \sigma}\mathrm{vec}(\Gamma^{-1} - \Gamma^{-1}D_t^{-1}(\lambda)\varepsilon_t(\varphi)\varepsilon'_t(\varphi)D_t^{-1}(\lambda)\Gamma^{-1}), \tag{S.9}$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \tilde{\lambda}\partial \tilde{\lambda}'} = -R_t^{(1)} - R_t^{(2)} - R_t^{(3)}, \tag{S.10}$$

$$
\begin{aligned}
R_t^{(1)} &= \frac{\partial \varepsilon_t'(\varphi)}{\partial \tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\frac{\partial \varepsilon_t(\varphi)}{\partial \tilde{\lambda}'},\\
R_t^{(2)} &= \frac{1}{4}\frac{\partial H_t'(\lambda)}{\partial \tilde{\lambda}}D_t^{-2}(\lambda)\Delta_t(\lambda)D_t^{-2}(\lambda)\frac{\partial H_t(\lambda)}{\partial \tilde{\lambda}'},\\
R_t^{(3)} &= (\varepsilon_t'(\varphi)\otimes I_N)\frac{\partial}{\partial \tilde{\lambda}'}\text{vec}\left[\frac{\partial \varepsilon_t'(\varphi)}{\partial \tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right]\\
&\quad + (\zeta_t'(\lambda)\otimes I_N)\frac{\partial}{\partial \tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial \tilde{\lambda}}D_t^{-2}(\lambda)\right]\\
&\quad - \frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial \tilde{\lambda}}D_t^{-2}(\lambda)[\tilde{\eta}_t(\lambda)\Gamma^{-1}D_t^{-1}(\lambda)+\tilde{\Delta}_t(\lambda)D_t^{-1}(\lambda)]\frac{\partial \varepsilon_t(\varphi)}{\partial \tilde{\lambda}'},\\
\frac{\partial^2 l_t(\lambda)}{\partial\varphi\partial\sigma'} &= \frac{\partial \varepsilon_t'(\varphi)}{\partial \varphi}(\varepsilon_t'(\varphi)D_t^{-1}(\lambda)\Gamma^{-1}\otimes D_t^{-1}(\lambda)\Gamma^{-1})\mathcal{K}\\
&\quad - \frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial \varphi}D_t^{-2}(\lambda)\frac{\partial \zeta_t(\lambda)}{\partial \sigma'}, \qquad \text{(S.11)}\\
\frac{\partial^2 l_t(\lambda)}{\partial\delta\partial\sigma'} &= -\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial \delta}D_t^{-2}(\lambda)\frac{\partial \zeta_t(\lambda)}{\partial \sigma'}, \qquad \text{(S.12)}\\
\frac{\partial \zeta_t(\lambda)}{\partial \sigma'} &= (\eta_t'(\lambda)\Gamma^{-1}\otimes\tilde{\eta}_t(\lambda))(I_m\otimes\Gamma^{-1})\mathcal{K}, \qquad \text{(S.13)}\\
\frac{\partial^2 l_t(\lambda)}{\partial\sigma\partial\sigma'} &= \frac{1}{2}\mathcal{K}'(\Gamma^{-1}\otimes I_m)\Big[I_{m^2}-(\Gamma^{-1}D_t^{-1}(\lambda)\varepsilon_t(\varphi)\varepsilon_t'(\varphi)D_t^{-1}(\lambda)\otimes I_m)\\
&\quad - (I_m\otimes\Gamma^{-1}D_t^{-1}(\lambda)\varepsilon_t(\varphi)\varepsilon_t'(\varphi)D_t^{-1}(\lambda))\Big](I_m\otimes\Gamma^{-1})\mathcal{K}, \qquad \text{(S.14)}
\end{aligned}
$$

where N is the length of $\tilde{\lambda}$, $\zeta_t = \Pi - \tilde{\eta}_t\Gamma^{-1}\eta_t, \Pi = (1,\ldots,1)'_{m\times 1}$, $\Delta_t = \tilde{\eta}_t\Gamma^{-1}\tilde{\eta}_t + \tilde{\Delta}_t\tilde{\eta}_t$, $\tilde{\Delta}_t = \text{diag}(e_1'\Gamma^{-1}\eta_t,\ldots,e_m'\Gamma^{-1}\eta_t)$, $e_i = (0,\ldots,0,1,0,\ldots,0)'$ of which the ith element is 1. $\tilde{\eta}_t = \text{diag}(\eta_{1t},\ldots,\eta_{mt})$ with $\eta_{it} = \varepsilon_{it}/\sqrt{h_{it}}$, $i = 1,\ldots,m$. First, we give four lemmas.

Lemma A.1 *Let $\xi_{\rho t}$ be defined as in Assumption 3.1. If Assumptions 2.1-2.4 hold, then exist constants C and $\rho\in(0,1)$, such that the following holds* *uniformly in* Θ*:*

(i) $\|\varepsilon_{t-1}(\varphi)\|$, $\left\|\frac{\partial\varepsilon_t(\varphi)}{\partial\varphi'}\right\|$, *and* $\left\|\frac{\partial^2\varepsilon_{it}(\varphi)}{\partial\varphi\partial\varphi'}\right\|$ *are bounded a.s. by* $C\xi_{\rho t-1}$ *for* $i = 1, \ldots, m$,

(ii) $\|H_t(\lambda)\|$ *is bounded a.s. by* $C\xi^2_{\rho t-1}$.

Proof. For (i), by (2.4), we have

$$\left\|(Y'_{t-i} - \mu') \otimes I_m\right\| = \sqrt{m}\,\|Y_{t-i} - \mu\|, \tag{S.15}$$

$$\sum_{i=1}^{p}\left\|\Psi^{-1}(L)[(Y'_{t-i} - \mu') \otimes I_m]\right\| = \sum_{i=1}^{p}\left\|\sum_{k=0}^{\infty}\gamma_k[(Y'_{t-i-k} - \mu') \otimes I_m]\right\|, \tag{S.16}$$

$$\sum_{i=1}^{q}\left\|\Psi^{-1}(L)(\varepsilon'_{t-i}(\varphi) \otimes I_m)\right\| = \sum_{i=1}^{q}\left\|\Psi^{-1}(L)[\Psi^{-1}(L)\Phi(L)(Y'_{t-i} - \mu') \otimes I_m]\right\|. \tag{S.17}$$

Using (2.4), (S.4), (S.15)-(S.17),

$$\sup_{\Theta}\|\varepsilon_{t-1}(\varphi)\| = \sup_{\Theta}\left\|\sum_{k=0}^{\infty}\gamma_{1k}(Y_{t-k-1} - \mu)\right\| \le C\left(1 + \sum_{k=0}^{\infty}\rho^k\|Y_{t-k-1}\|\right) = C\xi_{\rho t-1},$$

$$\begin{aligned}\sup_{\Theta}\left\|\frac{\partial\varepsilon_t(\varphi)}{\partial\varphi'}\right\| &\le \sup_{\Theta}\left\|\Psi^{-1}(1)\Phi(1)\right\| + \sup_{\Theta}\sum_{i=1}^{p}\left\|\Psi^{-1}(L)[(Y'_{t-i} - \mu') \otimes I_m]\right\| \\ &\quad + \sup_{\Theta}\sum_{i=1}^{q}\left\|\Psi^{-1}(L)[\varepsilon'_{t-i}(\varphi) \otimes I_m]\right\| \\ &\le C + C\sum_{i=1}^{p}\sum_{k=0}^{\infty}\rho^k\|Y_{t-i-k} - \mu\| + C\sum_{i=1}^{q}\sum_{k=0}^{\infty}\rho^k\|Y_{t-i-k} - \mu\| \le C\xi_{\rho t-1},\end{aligned}$$

$$\sup_{\Theta}\left\|\frac{\partial^2\varepsilon_{it}(\varphi)}{\partial\varphi\partial\varphi'}\right\| \le C\left(1+\sum_{i=0}^{\infty}\rho^i\|Y_{t-i-1}\|\right) = C\xi_{\rho t-1}.$$

Thus, (i) is proved.

For (ii), by (2.4) and (2.8),

$$\begin{aligned}
\sup_{\Theta}\|H_t(\lambda)\| &= \sup_{\Theta}\left\|\sum_{k=0}^{\infty}\Gamma_k W + \sum_{k=1}^{\infty}\Gamma_{1k}\vec{\varepsilon}_{t-k}(\varphi)\right\| \le C_1 + C_2\sum_{k=1}^{\infty}\rho^k\sup_{\Theta}\|\vec{\varepsilon}_{t-k}(\varphi)\| \\
&\le C_1 + C_3\sum_{k=1}^{\infty}\rho^k\sup_{\Theta}\|\varepsilon_{t-k}(\varphi)\|^2 = C_1 + C_3\sum_{k=1}^{\infty}\rho^k\sup_{\Theta}\left\|\sum_{l=0}^{\infty}\gamma_{1l}(Y_{t-k-l}-\mu)\right\|^2 \\
&\le C_4\left[1+\sum_{i=0}^{\infty}\rho^i\|Y_{t-i-1}\|\right]^2 = C_4\xi^2_{\rho t-1}.
\end{aligned}$$

Thus, (ii) holds. This completes the proof. □

Lemma A.2 *Let $\xi_{\rho t}$ be defined as in Assumption 3.1. Under Assumptions 2.1-2.4, there exists a neighborhood Θ_0 of λ_0 such that*

(i) $\sup_{\Theta_0}\left\|\frac{\partial H_t'(\lambda)}{\partial\delta}D_t^{-2}(\lambda)\right\| \le C\xi^{\tau_1}_{\rho t-1}$,

(ii) $\sup_{\Theta_0}\left\|\frac{1}{h_{it}(\lambda)}\frac{\partial^2 h_{it}(\lambda)}{\partial\delta\partial\delta'}\right\| \le C\xi^{\tau_1}_{\rho t-1}$,

for any $\tau_1\in(0,1)$, $\rho\in(0,1)$ and C is independent of τ_1 and t.

Proof. Rewrite (3.2) in matrix form as

$$\mathcal{H}_t = \underline{c}_t + G\mathcal{H}_{t-1}, \tag{S.18}$$

where G is defined as in (2.5),

$$\mathcal{H}_t = \begin{pmatrix} H_t(\lambda) \\ H_{t-1}(\lambda) \\ \vdots \\ H_{t-s+1}(\lambda) \end{pmatrix} \quad \text{and} \quad \underline{c}_t = \begin{pmatrix} W + \sum_{i=1}^{r} A_i \vec{\varepsilon}_{t-i}(\varphi) \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{S.19}$$

Setting $s_1 = m+(p+q)m^2$, $s_2 = 2m+(p+q)m^2+rm^2$, and $s_3 = 2m+(p+q)m^2+(r+s)m^2$, that is $s_3 = N$ in (S.10). By (2.7), (S.18), and (S.19),

$$\frac{\partial \mathcal{H}_t}{\partial \lambda_i} = \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{k} G^{j-1} G^{(i)} G^{k-j} \right\} \underline{c}_{t-k}, \qquad i = s_2+1, \ldots, s_3,$$

where $G^{(i)} = \partial G / \partial \lambda_i$ is a matrix whose entries are all 0, apart from a 1 located in the same place as λ_i in G. By abuse of convention, we denote $\mathcal{H}_{i_1 t}$ the i_1-th component of $\mathcal{H}_t$, $\underline{c}_{j_1,t-k}$ the j_1-th component of $\underline{c}_{t-k}$, and $G(i_1, j_1)$ the (i_1, j_1)-th entry of G. With similar argument to the proof of Theorem 10.9 in Francq and Zakoïan (2019), that is, $x/(1+x) \leq x^{\tau_1/2}$ for all $x \geq 0$ and $\tau_1 \in (0, 1]$, and the inequalities

$$\lambda_i \frac{\partial \mathcal{H}_t}{\partial \lambda_i} \leq \sum_{k=1}^{\infty} k G^k \underline{c}_{t-k}, \quad \lambda_i \frac{\partial \mathcal{H}_{i_1 t}}{\partial \lambda_i} \leq \sum_{k=1}^{\infty} k \sum_{j_1=1}^{m} G^k(i_1, j_1) \underline{c}_{j_1,t-k},$$

where $i = s_2+1, \ldots, s_3$, and setting $\omega = \inf_{1 \leq i \leq m} W_i$,

$$\mathcal{H}_{i_1 t} \geq \omega + \sum_{j_1=1}^{m} G^k(i_1, j_1) \underline{c}_{j_1,t-k}, \qquad \forall k.$$

By (2.7), we obtain

$$\begin{aligned}\frac{\lambda_i}{\mathcal{H}_{i_1t}}\frac{\partial\mathcal{H}_{i_1t}}{\partial\lambda_i} &\le \frac{\sum_{k=1}^{\infty}k\sum_{j_1=1}^{m}G^k(i_1,j_1)\underline{c}_{j_1,t-k}}{\omega+\sum_{j_1=1}^{m}G^k(i_1,j_1)\underline{c}_{j_1,t-k}} \\ &\le \sum_{k=1}^{\infty}k\left(\sum_{j_1=1}^{m}\frac{G^k(i_1,j_1)\underline{c}_{j_1,t-k}}{\omega}\right)^{\tau_1/2} \\ &\le C\sum_{k=1}^{\infty}k\left(\sum_{j_1=1}^{m}\rho_{j_1}^k\underline{c}_{j_1,t-k}\right)^{\tau_1/2}, \quad i=s_2+1,\ldots,s_3, \qquad \text{(S.20)}\end{aligned}$$

where the constant ρ_{j_1} (which also depends on i_1, τ_1) is on the interval $[0,1)$. Note that these inequalities are uniform on Θ_0 and can be extended to higher-order derivatives. As in the univariate case, by (S.20), for all $i_1=1,\ldots,m$, all $i=s_2+1,\ldots,s_3$,

$$\begin{aligned}&\sup_{\Theta_0}\left|\frac{1}{h_{i_1t}(\lambda)}\frac{\partial h_{i_1t}(\lambda)}{\partial\lambda_i}\right| \\ &\le C\sup_{\Theta_0}\sum_{k=1}^{\infty}k\left[\sum_{j_1=1}^{m}\rho_{j_1}^k\left(W_{j_1}+\sum_{l=1}^{r}\sum_{l_2=1}^{m}A_l(j_1,l_2)\varepsilon_{l_2,t-k-l}^2(\varphi)\right)\right]^{\tau_1/2} \\ &\le C\sup_{\Theta_0}\sum_{k=1}^{\infty}k\left[\sum_{j_1=1}^{m}\rho_{j_1}^kW_{j_1}+\sum_{j_1=1}^{m}\sum_{l=1}^{r}\sum_{l_2=1}^{m}\rho_{j_1}^kA_l(j_1,l_2)\varepsilon_{l_2,t-k-l}^2(\varphi)\right]^{\tau_1/2} \\ &\le C\sum_{k=1}^{\infty}k\left[\rho^k+\sum_{l=1}^{r}\rho^k\left\|\varepsilon_{t-k-l}(\varphi)\right\|^2\right]^{\tau_1/2} \\ &\le C\sum_{k=1}^{\infty}k\rho^k+C\sum_{k=1}^{\infty}k\sum_{l=1}^{r}\rho^k\left\|\varepsilon_{t-k-l}(\varphi)\right\|^{\tau_1} \\ &\le C\sum_{k=1}^{\infty}\rho^k+C\sum_{k=1}^{\infty}\rho^k\left\|\varepsilon_{t-k}(\varphi)\right\|^{\tau_1} \\ &\le C\left[1+\sum_{k=1}^{\infty}\rho^k\left\|\varepsilon_{t-k}(\varphi)\right\|^{\tau_1}\right], \qquad \text{(S.21)}\end{aligned}$$

where $A_l(j_1, l_2)$ is the (j_1, l_2)-th entry of matrix A_l, and constants $\rho \in (0,1)$ and $C > 0$ may vary across different places. Similarly, we can show that

$$\sup_{\Theta_0} \left| \frac{\lambda_i}{h_{i_1 t}(\lambda)} \frac{\partial h_{i_1 t}(\lambda)}{\partial \lambda_i} \right| \leq C \left[1 + \sum_{k=1}^{\infty} \rho^k \left\| \varepsilon_{t-k}(\varphi) \right\|^{\tau_1} \right], \qquad i = s_1 + 1, \ldots, s_2. \tag{S.22}$$

Let ρ be the maximizer of the ρ in Lemma A.1(i), (S.21), and (S.22), similar to (A.12) in Ling (2007), we can show,

$$\sum_{k=1}^{\infty} \rho^k \left\| \varepsilon_{t-k}(\varphi) \right\|^{\tau_1} \leq C \xi_{\tilde{\rho} t-1}^{\tau_1}, \tag{S.23}$$

for some $\tilde{\rho} \in (0,1)$. Thus, using (S.21)-(S.23), (i) holds. Using a similar method, (ii) holds. This completes the proof. □

Lemma A.3 *Let $\xi_{\rho t}$ be defined as in Assumption 3.1. Under Assumptions 2.1-2.4, there exist constants C and $\rho \in (0,1)$, such that*

(i) $\sup_{\Theta} \left\| \frac{\partial H_t'(\lambda)}{\partial \varphi} D_t^{-1}(\lambda) \right\| \leq C \xi_{\rho t-1}$,

(ii) $\sup_{\Theta} \left\| \frac{1}{\sqrt{h_{it}(\lambda)}} \frac{\partial^2 h_{it}(\lambda)}{\partial \varphi \partial \varphi'} \right\| \leq C \xi_{\rho t-1}$,

(iii) $\sup_{\Theta} \left\| \frac{1}{\sqrt{h_{it}(\lambda)}} \frac{\partial^2 h_{it}(\lambda)}{\partial \delta \partial \varphi'} \right\| \leq C \xi_{\rho t-1}$.

Proof. By (2.8) and (S.6),

$$H_t(\lambda) = \sum_{k=0}^{\infty} \Gamma_k W + \sum_{k=0}^{\infty} \Gamma_{1k} \vec{\varepsilon}_{t-k}(\varphi),$$

$$h_{i_1 t}(\lambda) \geq \Gamma_{1k}(i_1, j_1) \varepsilon_{j_1, t-k}^2(\varphi), \tag{S.24}$$

$$\frac{\partial H_t(\lambda)}{\partial \varphi'} = 2\sum_{k=1}^{\infty}\Gamma_{1k}\left(\bar{\varepsilon}^{*}_{t-k}(\varphi)\frac{\partial \varepsilon_{t-k}(\varphi)}{\partial \varphi'}\right),$$

$$\frac{\partial h_{i_1 t}(\lambda)}{\partial \varphi'} = 2\sum_{k=1}^{\infty}\sum_{j_1=1}^{m}\Gamma_{1k}(i_1,j_1)\varepsilon_{j_1,t-k}(\varphi)\frac{\partial \varepsilon_{j_1,t-k}(\varphi)}{\partial \varphi'}, \tag{S.25}$$

where $\Gamma_{1k}(i_1, j_1)$ is the (i_1, j_1)-th entry of matrix Γ_{1k}.

By (2.8), (S.24) and (S.25), and Lemma A.1(i),

$$\begin{aligned}
&\sup_{\Theta}\left\|\frac{\partial H_t'(\lambda)}{\partial \varphi}D_t^{-1}(\lambda)\right\| \\
&\leq C\sum_{i_1=1}^{m}\sup_{\Theta}\left\|\frac{\partial h_{i_1 t}(\lambda)}{\partial \varphi}\frac{1}{\sqrt{h_{it}(\lambda)}}\right\| \\
&= C\sum_{i_1=1}^{m}\sup_{\Theta}\left\|\frac{\partial h_{i_1 t}(\lambda)}{\partial \varphi'}\frac{1}{\sqrt{h_{it}(\lambda)}}\right\| \\
&\leq C\sum_{k=1}^{\infty}\sum_{i_1=1}^{m}\sum_{j_1=1}^{m}\sup_{\Theta}\left\|\frac{\Gamma_{1k}(i_1,j_1)\varepsilon_{j_1,t-k}(\varphi)}{\sqrt{h_{it}(\lambda)}}\frac{\partial \varepsilon_{j_1,t-k}(\varphi)}{\partial \varphi}\right\| \\
&\leq C\sum_{k=1}^{\infty}\sum_{i_1=1}^{m}\sum_{j_1=1}^{m}\sup_{\Theta}\left\|\sqrt{\Gamma_{1k}(i_1,j_1)}\frac{\partial \varepsilon_{j_1,t-k}(\varphi)}{\partial \varphi}\right\| \leq C\xi_{\rho t-1}.
\end{aligned}$$

Similarly, we can prove that (ii) and (iii) hold. This completes the proof. □

Lemma A.4 *Under Assumptions 2.1-2.4, for any $\iota_1 \in (0, 1)$, it follows that*

(i) $\sup_{\Theta}|l_t(\lambda) - l_t^{\epsilon}(\lambda)| = O(\rho^t)R_t^{1+\iota_1}$,

(ii) $\sup_{\Theta}\left\|\frac{\partial l_t(\lambda)}{\partial \lambda} - \frac{\partial l_t^{\epsilon}(\lambda)}{\partial \lambda}\right\| = O(\rho^t)R_t^{2+\iota_1}$,

(iii) $\sup_{\Theta}\left\|\frac{\partial^2 l_t(\lambda)}{\partial \lambda \partial \lambda'} - \frac{\partial^2 l_t^{\epsilon}(\lambda)}{\partial \lambda \partial \lambda'}\right\| = O(\rho^t)R_t^{3+\iota_1}$,

where $R_t = 1 + \sum_{i=0}^{t}\xi_{\rho t-i}^2$, and $\xi_{\rho t}$ is defined as in Assumption 3.1 for some $\rho \in (0, 1)$.

Proof. For convenience, the initial values Y_i are assumed to be zero when $i \leq 0$. When the initial values are not zero, the proof is similar. By (2.4),

$$\varepsilon_t(\varphi) = \sum_{k=0}^{\infty} \gamma_{1k}(Y_{t-k} - \mu), \quad \varepsilon_t^{\epsilon}(\varphi) = \sum_{k=0}^{t-1} \gamma_{1k}(Y_{t-k} - \mu),$$

$$\sup_{\Theta} \|\varepsilon_t(\varphi) - \varepsilon_t^{\epsilon}(\varphi)\| = O(1) \sum_{k=t}^{\infty} \rho^k \|Y_{t-k} - \mu\|$$
$$= O(\rho^t) \left[1 + \sum_{k=0}^{\infty} \rho^k \|Y_{-k}\|\right] = O(\rho^t)\xi_{\rho 0}. \tag{S.26}$$

By Lemma A.1 (i) and (S.26),

$$\begin{aligned}
\sup_{\Theta} \|\vec{\varepsilon}_t(\varphi) - \vec{\varepsilon}_t^{\epsilon}(\varphi)\| &\leq C \sup_{\Theta} \sum_{i=1}^{m} |\varepsilon_{it}^2(\varphi) - \varepsilon_{it}^{\epsilon 2}(\varphi)| \\
&\leq C \sup_{\Theta} \left[\sum_{i=1}^{m} |\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi)|\right] \sup_{\Theta} \left[\sum_{i=1}^{m} |\varepsilon_{it}(\varphi) + \varepsilon_{it}^{\epsilon}(\varphi)|\right] \\
&\leq C \sup_{\Theta} \|\varepsilon_t(\varphi) - \varepsilon_t^{\epsilon}(\varphi)\| \sup_{\Theta} \|\varepsilon_t(\varphi) + \varepsilon_t^{\epsilon}(\varphi)\| \\
&= O(\rho^t)\xi_{\rho 0}\xi_{\rho t}. \qquad (S.27)
\end{aligned}$$

By (2.8),

$$H_t(\lambda) = \sum_{k=0}^{\infty} \Gamma_k W + \sum_{k=0}^{\infty} \Gamma_{1k}\vec{\varepsilon}_{t-k}(\varphi), \tag{S.28}$$

$$H_t^{\epsilon}(\lambda) = \sum_{k=0}^{t-1} \Gamma_k W + \sum_{k=0}^{t-1} \Gamma_{1k}\vec{\varepsilon}_{t-k}^{\epsilon}(\varphi). \tag{S.29}$$

By Lemma A.1(i), (S.27)-(S.29),

$$\sup_{\Theta} \|H_t(\lambda) - H_t^{\epsilon}(\lambda)\|$$
$$\leq C \sum_{k=0}^{t-1} \rho^k \sup_{\Theta} \left\|\vec{\varepsilon}_{t-k}(\varphi) - \vec{\varepsilon}_{t-k}^{\epsilon}(\varphi)\right\| + C \sum_{k=t}^{\infty} \rho^k \sup_{\Theta} \|\vec{\varepsilon}_{t-k}(\varphi)\|$$

$$\leq C \sum_{k=0}^{t-1} \rho^t \xi_{\rho 0} \xi_{\rho t-k} + C \sum_{k=t}^{\infty} \rho^k \left\| \vec{\varepsilon}_{t-k}(\varphi) \right\|$$

$$\leq C\rho^t \xi_{\rho 0}^2 + C\rho^t \xi_{\rho 0} \sum_{k=0}^{t-1} \xi_{\rho t-k} = O(\rho^t) R_t. \tag{S.30}$$

Thus, by (S.26)-(S.30),

$$\sup_{\Theta} |h_{it}(\lambda) - h_{it}^{\epsilon}(\lambda)| \leq \sup_{\Theta} \|H_t(\lambda) - H_t^{\epsilon}(\lambda)\| = O(\rho^t) R_t, \tag{S.31}$$

$$\sup_{\Theta} |\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi)| \leq \sup_{\Theta} \|\varepsilon_t(\varphi) - \varepsilon_t^{\epsilon}(\varphi)\| = O(\rho^t) \xi_{\rho 0}, \tag{S.32}$$

$$\begin{aligned} \sup_{\Theta} (\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi))^2 &\leq C \sup_{\Theta} \|\varepsilon_t(\varphi) - \varepsilon_t^{\epsilon}(\varphi)\| \sup_{\Theta} [\|\varepsilon_t(\varphi)\| + \|\varepsilon_t^{\epsilon}(\varphi)\|] \\ &= O(\rho^t) \xi_{\rho 0} \xi_{\rho t}, \end{aligned} \tag{S.33}$$

where $i = 1, \ldots, m$, $0 < \rho < 1$. Because $h_{it}(\lambda)$ and $h_{it}^{\epsilon}(\lambda)$ have lower bounds uniformly in all t, i and λ, by (S.31),

$$\begin{aligned} \sup_{\Theta} \left| \frac{1}{\sqrt{h_{it}(\lambda)}} - \frac{1}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right| &\leq C \sup_{\Theta} \left| \frac{1}{\sqrt{h_{it}(\lambda)}} - \frac{1}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^{\iota_1} \\ &= C \sup_{\Theta} \left| \frac{\sqrt{h_{it}(\lambda)} - \sqrt{h_{it}^{\epsilon}(\lambda)}}{\sqrt{h_{it}(\lambda)}\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^{\iota_1} \\ &\leq C \sup_{\Theta} |h_{it}(\lambda) - h_{it}^{\epsilon}(\lambda)|^{\iota_1} \\ &= O(\rho^t) R_t^{\iota_1}, \end{aligned} \tag{S.34}$$

$$\begin{aligned} \sum_{i=1}^{m} \left| \frac{\varepsilon_{it}(\varphi)}{\sqrt{h_{it}(\lambda)}} - \frac{\varepsilon_{it}^{\epsilon}(\varphi)}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^2 = O(1) \sum_{i=1}^{m} \Bigg[\varepsilon_{it}^2(\varphi) \left| \frac{1}{\sqrt{h_{it}(\lambda)}} - \frac{1}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^2 \\ + \left(\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi)\right)^2 \Bigg], \end{aligned} \tag{S.35}$$

for any $\iota_1 \in (0,1)$. By (S.31)-(S.35), we have

$$\begin{aligned}
&\sup_{\Theta} |\ln|D_t(\lambda)\Gamma D_t(\lambda)| - \ln|D_t^{\epsilon}(\lambda)\Gamma D_t^{\epsilon}(\lambda)|| \\
&\quad \le \sum_{i=1}^{m} \sup_{\Theta} \left| \ln \frac{h_{it}(\lambda)}{h_{it}^{\epsilon}(\lambda)} \right| \le \sum_{i=1}^{m} \sup_{\Theta} \left| \frac{h_{it}(\lambda) - h_{it}^{\epsilon}(\lambda)}{h_{it}^{\epsilon}(\lambda)} \right| \\
&\quad \le C \sum_{i=1}^{m} \sup_{\Theta} |h_{it}(\lambda) - h_{it}^{\epsilon}(\lambda)| = O(\rho^t) R_t, \qquad \text{(S.36)}
\end{aligned}$$

$$\begin{aligned}
&\sup_{\Theta} |\varepsilon_t'(\varphi)(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_t(\varphi) - \varepsilon_t^{\epsilon'}(\varphi)(D_t^{\epsilon}(\lambda)\Gamma D_t^{\epsilon}(\lambda))^{-1}\varepsilon_t^{\epsilon}(\varphi)| \\
&\quad = \sup_{\Theta} |2\varepsilon_t'(\varphi) D_t^{-1}(\lambda)\Gamma^{-1}(D_t^{-1}(\lambda)\varepsilon_t(\varphi) - D_t^{\epsilon-1}(\lambda)\varepsilon_t^{\epsilon}(\varphi)) \\
&\qquad - (\varepsilon_t'(\varphi) D_t^{-1}(\lambda) - \varepsilon_t^{\epsilon'}(\varphi) D_t^{\epsilon-1}(\lambda))\Gamma^{-1}(D_t^{-1}(\lambda)\varepsilon_t(\varphi) - D_t^{\epsilon-1}(\lambda)\varepsilon_t^{\epsilon}(\varphi))| \\
&\quad \le C \sup_{\Theta} \|\varepsilon_t(\varphi)\| \left(\sum_{i=1}^{m} \left| \frac{\varepsilon_{it}(\varphi)}{\sqrt{h_{it}(\lambda)}} - \frac{\varepsilon_{it}^{\epsilon}(\varphi)}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^2 \right)^{1/2} \\
&\qquad + C \sup_{\Theta} \sum_{i=1}^{m} \left| \frac{\varepsilon_{it}(\varphi)}{\sqrt{h_{it}(\lambda)}} - \frac{\varepsilon_{it}^{\epsilon}(\varphi)}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^2 \\
&\quad \le C \sup_{\Theta} \sum_{i=1}^{m} \Bigg[\|\varepsilon_t(\varphi)\| \, |\varepsilon_{it}(\varphi)| \left| \frac{1}{\sqrt{h_{it}(\lambda)}} - \frac{1}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right| \\
&\qquad\qquad + \varepsilon_{it}^2(\varphi) \left| \frac{1}{\sqrt{h_{it}(\lambda)}} - \frac{1}{\sqrt{h_{it}^{\epsilon}(\lambda)}} \right|^2 \Bigg] \\
&\qquad + C \sup_{\Theta} \sum_{i=1}^{m} \left[\|\varepsilon_t(\varphi)\| \, |\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi)| + (\varepsilon_{it}(\varphi) - \varepsilon_{it}^{\epsilon}(\varphi))^2 \right] \\
&\quad = O(1)\xi_{\rho t}^2 O(\rho^t) R_t^{\iota_1} + O(1)\xi_{\rho t}^2 O(\rho^t) R_t^{\iota_1} + O(1)\xi_{\rho t} O(\rho^t)\xi_{\rho 0} + O(\rho^t)\xi_{\rho t}\xi_{\rho 0} \\
&\quad = O(\rho^t) R_t^{1+\iota_1}. \qquad \text{(S.37)}
\end{aligned}$$

Thus, by (S.36)-(S.37), we can show that (i) holds. Similarly, we can show that (ii) and (iii) hold. This completes the proof. □

Proof of Theorem 3.1(i)

Proof. First, the parameter space Θ is compact and λ_0 is an interior point in Θ. Second, $L_{sn}^{\epsilon}(\lambda)$ is continuous in Θ and is a measurable function of $\{Y_t, t = n, n-1, \cdots\}$ for all $\lambda \in \Theta$. Third, by Assumption 2.3, $\|D_t(\lambda)\Gamma D_t(\lambda)\|$ has a lower bound uniformly over Θ. By Lemma A.1,

$$\begin{aligned}\sup_{\Theta} \|\varepsilon_t(\varphi)\| &\leq \|\varepsilon_{0t}\| + \sup_{\Theta} \|\varphi - \varphi_0\| \sup_{\Theta} \left\| \frac{\partial \varepsilon_t(\varphi)}{\partial \varphi'} \right\| \\ &\leq \|D_{0t}\| \|\eta_{0t}\| + C\xi_{\rho t-1} \leq C \|H_{0t}\|^{1/2} \|\eta_{0t}\| + C\xi_{\rho t-1} \\ &\leq C\xi_{\rho t-1}(1 + \|\eta_{0t}\|). \qquad \text{(S.38)}\end{aligned}$$

By (S.38), Assumption 3.1, and Lemma A.1(ii),

$$\begin{aligned}E\sup_{\Theta} |w_t\varepsilon_t'(\varphi)(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_t(\varphi)| &\leq E\sup_{\Theta} w_t \|\varepsilon_t(\varphi)\|^2 \left\|(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right\| \\ &\leq CEw_t\xi_{\rho t-1}^2 E(1 + \|\eta_{0t}\|)^2 < \infty, \\ E\sup_{\Theta} \left|w_t \ln \left|D_t(\lambda)\Gamma D_t(\lambda)\right|\right| &\leq CE\sup_{\Theta} \left|w_t \ln \prod_{i=1}^m h_{it}(\lambda)\right| \\ &\leq CE\sup_{\Theta} \left[w_t \sum_{i=1}^m |h_{it}(\lambda) - 1|\right] \\ &\leq C + CE\sup_{\Theta} w_t \sum_{i=1}^m \left|h_{it}(\lambda)\right| \\ &\leq C + CE\sup_{\Theta} w_t \|H_t(\lambda)\| \\ &\leq C + CEw_t\xi_{\rho t-1}^2 < \infty.\end{aligned}$$

Thus, we can claim that $E\sup_{\Theta} |w_t l_t(\lambda)| < \infty$.

By Theorem 3.1 in Ling and McAleer (2003),

$$\sup_{\Theta} |L_{sn}(\lambda) - E[w_t l_t(\lambda)]| \xrightarrow{p} 0. \tag{S.39}$$

It's straightforward to show that $\sup_{\Theta}[|l_t(\lambda)| + |l_t^{\epsilon}(\lambda)|] < C\xi_{\rho t}^2$. By Assumptions 3.1-3.2 and Lemma A.4(i),

$$\begin{aligned}\sup_{\Theta} |w_t l_t(\lambda) - w_t^{\epsilon} l_t^{\epsilon}(\lambda)| &\leq w_t \sup_{\Theta} |l_t(\lambda) - l_t^{\epsilon}(\lambda)| + |w_t - w_t^{\epsilon}| \sup_{\Theta} |l_t^{\epsilon}(\lambda)| \\ &= O(\rho^t) R_t^{1+\iota_1} + O(1)|w_t - w_t^{\epsilon}|\xi_{\rho t}^2 \equiv A_{1t} + A_{2t},\end{aligned} \tag{S.40}$$

$$E(|w_t - w_t^{\epsilon}|\xi_{\rho t}^2)^{\iota_0/8} \leq \left[E|w_t - w_t^{\epsilon}|^{\iota_0/4} E\xi_{\rho t}^{\iota_0/2}\right]^{1/2} = O(t^{-1}), \tag{S.41}$$

where ι_0 is the same as in Assumption 3.2, and ι_1 is the same as in Lemma A.4. To prove $\sum_{t=1}^n A_{1t}/n = o_p(1)$, it is sufficient to show

$$E\left|\frac{1}{n}\sum_{t=1}^n A_{1t}\right|^{\iota/(1+\iota_1)} \to 0, \quad \text{as} \quad n \to \infty,$$

where ι is the same as in Assumption 2.5. By Assumption 2.5,

$$\begin{aligned}E\left|\frac{1}{n}\sum_{t=1}^n A_{1t}\right|^{\iota/(1+\iota_1)} &\leq \frac{1}{n^{\iota/(1+\iota_1)}} \sum_{t=1}^n E|A_{1t}|^{\iota/(1+\iota_1)} \\ &= \frac{1}{n^{\iota/(1+\iota_1)}} \sum_{t=1}^n E\left|O(\rho^t) R_t^{1+\iota_1}\right|^{\iota/(1+\iota_1)} \\ &= \frac{1}{n^{\iota/(1+\iota_1)}} \sum_{t=1}^n O(\rho_1^t) E R_t^{\iota} \\ &= \frac{1}{n^{\iota/(1+\iota_1)}} \sum_{t=1}^n O(\rho_1^t) E\left[1 + \sum_{i=0}^t \xi_{\rho t-i}^2\right]^{\iota}\end{aligned}$$

$$= \frac{1}{n^{\iota/(1+\iota_1)}} \sum_{t=1}^{n} O(\rho_1^t) E\left[1 + \sum_{i=0}^{t}\left(1 + \sum_{j=0}^{\infty} \rho^j |y_{t-i-j}|\right)^2\right]^{\iota}$$

$$\to 0. \tag{S.42}$$

Similarly, by (S.41),

$$E\left[\frac{1}{n}\sum_{t=1}^{n} A_{2t}\right]^{\iota_0/8} \leq \frac{1}{n^{\iota_0/8}} \sum_{t=1}^{n} E\big[A_{2t}^{\iota_0/8}\big] \leq \frac{1}{n^{\iota_0/8}} \sum_{t=1}^{n} O(t^{-1}) \to 0. \tag{S.43}$$

By (S.40), (S.42)-(S.43), we can show that

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\Theta} |w_t l_t(\lambda) - w_t^{\epsilon} l_t^{\epsilon}(\lambda)| = o_p(1). \tag{S.44}$$

By (S.39) and (S.44), it follows that

$$\sup_{\Theta} |L_{sn}^{\epsilon}(\lambda) - E[w_t l_t(\lambda)]| = o_p(1). \tag{S.45}$$

Fourth, we need to prove $E[w_t l_t(\lambda)]$ achieves its maximum at $\lambda = \lambda_0$.

$$\begin{aligned} E[w_t l_t(\lambda)] &= E[w_t(-\ln|D_t(\lambda)\Gamma D_t(\lambda)| - \varepsilon_t'(\varphi)(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_t(\varphi))] \\ &= \{-E[w_t \ln|D_t(\lambda)\Gamma D_t(\lambda)|] - E[w_t \varepsilon_{0t}'(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\varepsilon_{0t}]\} \\ &\quad - E\left[w_t(\varepsilon_t(\varphi) - \varepsilon_{0t})'(D_t(\lambda)\Gamma D_t(\lambda))^{-1}(\varepsilon_t(\varphi) - \varepsilon_{0t})\right] \\ &\equiv L_1(\lambda) + L_2(\lambda). \end{aligned}$$

If and only if $\varepsilon_t(\varphi) = \varepsilon_{0t}$, $L_2(\lambda)$ obtains its maximum at zero. According to Lemma 4.2 in Ling and McAleer (2003),

$$\varepsilon_t(\varphi) - \varepsilon_{0t} = \frac{\partial \varepsilon_t}{\partial \varphi'}\Big|_{\varphi*}(\varphi - \varphi_0) = 0,$$

where $\varphi*$ lies between φ and φ_0, and the above equation holds if and only if $\varphi = \varphi_0$.

Let $M_t = (D_t(\lambda)\Gamma D_t(\lambda))^{-1/2}(D_{0t}\Gamma_0 D_{0t})(D_t(\lambda)\Gamma D_t(\lambda))^{-1/2}$,

$$\begin{aligned} L_1(\lambda) &= E\big\{w_t[-\log|D_t(\lambda)\Gamma D_t(\lambda)| - trace(M_t)]\big\} \\ &= Ew_t[\log|M_t| - trace(M_t))] - Ew_t \log|D_{0t}\Gamma_0 D_{0t}|]. \end{aligned}$$

Similar to Lemma 4.4 in Ling and McAleer (2003),

$$\log|M_t| - trace(M_t) \leq -m,$$

the equality holds only if $M_t = I_m$ with probability one. Thus, $L_1(\lambda)$ reaches its maximum $-mEw_t - Ew_t \log|D_{0t}\Gamma D_{0t}|$ if and only if $D_t(\lambda)\Gamma D_t(\lambda) = D_{0t}\Gamma_0 D_{0t}$. By the definition of Γ_0, we have $h_{0it} = h_{it}$, and $\Gamma = \Gamma_0$. Since $\max_{\lambda\in\Theta} L(\lambda) \leq \max_{\lambda\in\Theta} L_1(\lambda) + \max_{\lambda\in\Theta} L_2(\lambda)$, then $\max_{\lambda\in\Theta} L(\lambda) = -mEw_t - Ew_t \log|D_{0t}\Gamma_0 D_{0t}|$ if and only if $\max_{\lambda\in\Theta} L_2(\lambda) = 0$ and $\max_{\lambda\in\Theta} L_1(\lambda) = -mEw_t - Ew_t \log|D_{0t}\Gamma_0 D_{0t}|$, which occurs if and only if $\varphi = \varphi_0$, $\Gamma = \Gamma_0$, $h_{it} = h_{0it}$. From $\varphi = \varphi_0$ and $h_{it} = h_{0it}$, we have

$$(H_t - H_{0t})\big|_{\varphi=\varphi_0} = \frac{\partial H_t}{\partial \delta'}\Big|_{(\varphi_0,\delta^*)}(\delta - \delta_0) = 0$$

with probability one, where δ^* lies between δ and δ_0. Again, according to Lemma 4.2 in Ling and McAleer (2003), the above equality holds if and only if $\delta = \delta_0$. Thus, $L(\lambda)$ is uniquely maximized at λ_0. Thus, all the conditions in Theorem 4.1.1 in Amemiya (1985) have been established. This completes the proof. □

Proof of Theorem 3.1(ii)

Proof. First, $\hat{\lambda}_{sn} \xrightarrow{p} \lambda_0$ as $n \to \infty$. Second, $\partial^2 l_t(\lambda)/\partial\lambda\partial\lambda'$ exists and is continuous in Θ. Third, since $\|D_t(\lambda)\|$ has a lower bound uniformly in Θ, by (S.38),

$$\begin{aligned}\sup_{\Theta}\|\Delta_t\| &= \sup_{\Theta}\left\|\tilde{\eta}_t(\lambda)\Gamma^{-1}\tilde{\eta}_t(\lambda) + \tilde{\Delta}_t\tilde{\eta}_t(\lambda)\right\| \\ &\leq C\sup_{\Theta}\|\eta_t(\lambda)\|^2 \leq C\sup_{\Theta}\|\varepsilon_t(\varphi)\|^2 \\ &\leq C(1+\|\eta_{0t}\|)^2\xi^2_{\rho t-1}, \end{aligned} \tag{S.46}$$

$$\begin{aligned}&\sup_{\Theta}\left\|\tilde{\eta}_t(\lambda)\Gamma^{-1}D_t^{-1}(\lambda) + \tilde{\Delta}_t D_t^{-1}(\lambda)\right\| \\ &\qquad\leq \sup_{\Theta}\|\tilde{\eta}_t(\lambda)\|\left\|\Gamma^{-1}\right\|\left\|D_t^{-1}(\lambda)\right\| + \sup_{\Theta}\left\|\tilde{\Delta}_t\right\|\left\|D_t^{-1}(\lambda)\right\| \\ &\qquad\leq C\sup_{\Theta}\|\eta_t(\lambda)\| \leq C(1+\|\eta_{0t}\|)\xi_{\rho t-1}, \end{aligned} \tag{S.47}$$

$$\sup_{\Theta}\|\varepsilon_t'(\varphi)\otimes I_N\| \leq C\sup_{\Theta}\|\varepsilon_t(\varphi)\| \leq C(1+\|\eta_{0t}\|)\xi_{\rho t-1}, \tag{S.48}$$

$$\sup_{\Theta}\|\zeta_t'(\lambda)\otimes I_N\| \leq C\sup_{\Theta}\|\eta_t(\lambda)\|^2 \leq C(1+\|\eta_{0t}\|)^2\xi^2_{\rho t-1}. \tag{S.49}$$

By (S.10), (S.46), and Lemmas A.1-A.3, there exists a neighborhood Θ_0 of λ_0 such that

$$\sup_{\Theta_0}\left\|R_t^{(1)}\right\| \leq C\sup_{\Theta_0}\left\|\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right\|^2 \leq C\xi^2_{\rho t-1}, \tag{S.50}$$

$$\sup_{\Theta_0}\left\|R_t^{(2)}\right\| \leq C\sup_{\Theta_0}\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right\|^2 \sup_{\Theta_0}\|\Delta_t\| \leq C(1+\|\eta_{0t}\|)^2\xi_{\rho t-1}^4. \tag{S.51}$$

It is not hard to see

$$\begin{aligned}
&\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right] \\
&\quad = [(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\otimes I_N]\begin{pmatrix}\dfrac{\partial^2\varepsilon_{1t}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{\partial^2\varepsilon_{mt}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\end{pmatrix} \\
&\qquad + \left[I_m\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right](I_m\otimes D_t^{-1}(\lambda)\Gamma^{-1})\left[-\frac{1}{2}\text{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right] \\
&\qquad + \left[I_m\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right](D_t^{-1}(\lambda)\Gamma^{-1}\otimes I_m)\left[-\frac{1}{2}\text{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right],
\end{aligned} \tag{S.52}$$

$$\begin{aligned}
&\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right] \\
&\quad = \frac{1}{2}\begin{pmatrix}\dfrac{1}{h_{1t}(\lambda)}\dfrac{\partial^2 h_{1t}(\lambda)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{1}{h_{mt}(\lambda)}\dfrac{\partial^2 h_{mt}(\lambda)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\end{pmatrix} - \frac{1}{2}\begin{pmatrix}\dfrac{1}{h_{1t}^2(\lambda)}\dfrac{\partial h_{1t}(\lambda)}{\partial\tilde{\lambda}}\dfrac{\partial h_{1t}(\lambda)}{\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{1}{h_{mt}^2(\lambda)}\dfrac{\partial h_{mt}(\lambda)}{\partial\tilde{\lambda}}\dfrac{\partial h_{mt}(\lambda)}{\partial\tilde{\lambda}'}\end{pmatrix}.
\end{aligned} \tag{S.53}$$

By Lemmas A.1-A.3, (S.52)-(S.53), there exists a neighborhood Θ_0 of λ_0 such that

$$\sup_{\Theta_0}\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right]\right\|$$

$$\leq C\sup_{\Theta_0}\left\|(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\otimes I_N\right\|\sum_{i=1}^{m}\sup_{\Theta_0}\left\|\frac{\partial^2\varepsilon_{it}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\right\|$$

$$+C\sup_{\Theta_0}\left\|I_m\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right\|\sup_{\Theta_0}\left\|I_m\otimes D_t^{-1}(\lambda)\Gamma^{-1}\right\|$$

$$\times\sup_{\Theta_0}\left\|\text{diag}(e_1,\cdots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right\|$$

$$\leq C\xi_{\rho t-1}^2, \tag{S.54}$$

$$\sup_{\Theta_0}\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left(\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right)\right\|$$

$$\leq C\sum_{i=1}^{m}\sup_{\Theta_0}\left\|\frac{1}{h_{it}(\lambda)}\frac{\partial^2 h_{it}(\lambda)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\right\|$$

$$+C\sum_{i=1}^{m}\sup_{\Theta_0}\left\|\frac{1}{h_{it}^2(\lambda)}\frac{\partial h_{it}(\lambda)}{\partial\tilde{\lambda}}\frac{\partial h_{it}(\lambda)}{\partial\tilde{\lambda}'}\right\|$$

$$\leq C\xi_{\rho t-1}^2. \tag{S.55}$$

Denote $R_t^{(3)}\equiv R_{1t}^{(3)}+R_{2t}^{(3)}+R_{3t}^{(3)}$, by (S.47)-(S.49),(S.54)-(S.55), Lemmas A.1-A.3,

$$\sup_{\Theta}\left\|R_{1t}^{(3)}\right\|\leq\sup_{\Theta}\left\|\varepsilon_t'(\varphi)\otimes I_N\right\|\sup_{\Theta}\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right]\right\|$$

$$\leq C(1+\|\eta_{0t}\|)\xi_{\rho t-1}^3, \tag{S.56}$$

$$\sup_{\Theta_0}\left\|R_{2t}^{(3)}\right\|\leq\sup_{\Theta_0}\left\|\zeta'(\lambda)\otimes I_N\right\|\sup_{\Theta_0}\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left(\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right)\right\|$$

$$\leq C(1+\|\eta_{0t}\|)^2\xi_{\rho t-1}^4, \tag{S.57}$$

$$\sup_{\Theta_0}\left\|R_{3t}^{(3)}\right\|\leq C\sup_{\Theta_0}\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right\|\sup_{\Theta_0}\left\|\frac{\partial\varepsilon_t(\varphi)}{\partial\tilde{\lambda}'}\right\|$$

$$\times\sup_{\Theta_0}\left\|\tilde{\eta}_t(\lambda)\Gamma^{-1}D_t^{-1}(\lambda)+\tilde{\Delta}_t D_t^{-1}(\lambda)\right\|$$

$$\leq C(1 + \|\eta_{0t}\|)\xi^3_{\rho t-1}. \tag{S.58}$$

Thus, by (S.56)-(S.58),

$$\sup_{\Theta_0} \left\| R^{(3)}_t \right\| \leq C(1 + \|\eta_{0t}\|)^2 \xi^4_{\rho t-1}. \tag{S.59}$$

By (S.10), (S.50)-(S.51), and (S.59),

$$\begin{aligned}
\sup_{\Theta_0} \left\| \frac{\partial^2 l_t(\lambda)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'} \right\| &\leq \sup_{\Theta_0} \left\| R^{(1)}_t \right\| + \sup_{\Theta_0} \left\| R^{(2)}_t \right\| + \sup_{\Theta_0} \left\| R^{(3)}_t \right\| \\
&\leq C(1 + \|\eta_{0t}\|)^2 \xi^4_{\rho t-1}. \qquad (S.60)
\end{aligned}$$

By Assumption 3.1, we can show that

$$E \sup_{\Theta_0} \left\| \omega_t \frac{\partial^2 l_t(\lambda)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'} \right\| < \infty. \tag{S.61}$$

Similarly, we can show that

$$\begin{aligned}
E \sup_{\Theta} \left\| \omega_t \frac{\partial^2 l_t(\lambda)}{\partial\varphi\partial\sigma'} \right\| &< \infty, \\
E \sup_{\Theta_0} \left\| \omega_t \frac{\partial^2 l_t(\lambda)}{\partial\delta\partial\sigma'} \right\| &< \infty, \\
E \sup_{\Theta} \left\| \omega_t \frac{\partial^2 l_t(\lambda)}{\partial\sigma\partial\sigma'} \right\| &< \infty.
\end{aligned}$$

Thus, we can claim

$$E \sup_{\Theta_0} \left\| \omega_t \frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda'} \right\| < \infty.$$

By the ergodic theorem and Theorem 3.1 in Ling and McAleer (2003), we can show that $\partial^2 L_{sn}(\lambda)/\partial\lambda\partial\lambda'$ converges to $E\left[w_t \partial^2 l_t(\lambda)/\partial\lambda\partial\lambda'\right]$ uniformly in Θ_0 in probability. Similar to (S.44), using Lemma A.4 (iii), we can show that

$$\sup_{\Theta_0}\left\|\frac{\partial^2 L_{sn}(\lambda)}{\partial\lambda\partial\lambda'}-\frac{\partial^2 L^{\epsilon}_{sn}(\lambda)}{\partial\lambda\partial\lambda'}\right\|=o_p(1).$$

For any sequence of λ_n such that $\lambda_n\to\lambda_0$ in probability, by dominated convergence theorem, we can show that

$$\frac{\partial^2 L^{\epsilon}_{sn}(\lambda_n)}{\partial\lambda\partial\lambda'}=\Sigma_0+o_p(1). \tag{S.62}$$

Fourth, for the previous neighborhood Θ_0, (S.7)-(S.9), (S.38), (S.49), and Lemmas A.1-A.3,

$$\begin{aligned}
\sup_{\Theta}\left\|\frac{\partial l_t(\lambda)}{\partial\varphi}\right\| &\le \sup_{\Theta}\left\|\frac{\partial\varepsilon_t'(\varphi)}{\partial\varphi}\right\|\sup_{\Theta}\left\|(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right\|\sup_{\Theta}\|\varepsilon_t(\varphi)\| \\
&\quad+\frac{1}{2}\sup_{\Theta}\left\|\frac{\partial H_t'(\lambda)}{\partial\varphi}D_t^{-2}(\lambda)\right\|\sup_{\Theta}\|\zeta_t(\lambda)\| \\
&\le C\left[(1+\|\eta_{0t}\|)\xi^2_{\rho t-1}+(1+\|\eta_{0t}\|)^2\xi^3_{\rho t-1}\right] \\
&\le C(1+\|\eta_{0t}\|)^2\xi^3_{\rho t-1}. \\
\sup_{\Theta_0}\left\|\frac{\partial l_t(\lambda)}{\partial\delta}\right\| &\le C\sup_{\Theta}\left\|\frac{\partial H_t'(\lambda)}{\partial\delta}D_t^{-2}(\lambda)\right\|\sup_{\Theta}\|\zeta_t(\lambda)\| \\
&\le C(1+\|\eta_{0t}\|)^2\xi^3_{\rho t-1}, \\
\sup_{\Theta}\left\|\frac{\partial l_t(\lambda)}{\partial\sigma}\right\| &\le C\left\|\frac{\partial \mathrm{vec}(\Gamma)'}{\partial\sigma}\right\|\sup_{\Theta}\left\|\mathrm{vec}(\Gamma^{-1})\right\| \\
&\quad+C\left\|\frac{\partial \mathrm{vec}(\Gamma)'}{\partial\sigma}\right\|\sup_{\Theta}\left\|\mathrm{vec}(\Gamma^{-1}D_t^{-1}\varepsilon_t(\varphi)\varepsilon_t'(\varphi)D_t^{-1}\Gamma^{-1})\right\| \\
&\le C(1+\|\eta_{0t}\|)^2\xi^2_{\rho t-1}.
\end{aligned}$$

Thus, we can prove

$$\Omega_0 = E\left[w_t^2 \frac{\partial l_t(\lambda)}{\partial \lambda}\frac{\partial l_t(\lambda)}{\partial \lambda'}\right] < \infty.$$

Similarly to the proof of Lemma 4.2 in Ling and McAleer (2003), we can show $-\Sigma_0$ and Ω_0 are positive definite. By central limiting theorem, we have $\partial L_{sn}(\lambda_0)/\partial\lambda \xrightarrow{\mathcal{L}} N(0,\Omega_0)$. Similarly, we can show that $\sqrt{n}\|\partial L_{sn}(\lambda_0)/\partial\lambda - \partial L_{sn}^{\epsilon}(\lambda_0)/\partial\lambda\| = o_p(1)$ and hence $\partial L_{sn}^{\epsilon}(\lambda_0)/\partial\lambda \xrightarrow{\mathcal{L}} N(0,\Omega_0)$. Thus, all the conditions in Theorem 4.1.3 in Amemiya (1985) have been established and hence $\sqrt{n}(\lambda_{sn}-\lambda_0) \xrightarrow{\mathcal{L}} N(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1})$. This completes the proof. □

The following lemmas are the key results for Theorem 4.1.

Lemma A.5 *Let $\xi_{\rho t}$ be defined as in Assumption 3.1 and $\xi_{0\varrho t} = 1 + \sum_{i=0}^{\infty}\varrho^i \|\varepsilon_{0,t-i}\|$. If Assumptions 2.1-2.4 hold, then for any $\rho,\varrho \in (0,1)$, there exist constants $\varrho_1 \in (0,1), \tilde{\iota} \in (0,1/2)$ and C not depending on t such that*

(i) $\xi_{\rho t} \le C\xi_{0\varrho_1 t}$ *a.s.,*

(ii) $\dfrac{\xi_{0\varrho t-1}}{\sqrt{h_{0kt}}} \le C\xi_{0\varrho_1 t-1}^{1-\tilde{\iota}}$ *a.s. for $k = 1,\ldots,m$.*

Proof. By Assumptions 2.1-2.2, Y_t has the following expansion:

$$Y_t = \mu + \sum_{k=0}^{\infty}\gamma_{1k}\varepsilon_{0,t-k},$$

where γ_{1k} is defined as in (2.4). Thus, (i) can be directly proved. Let G_0

be defined as G in (2.5) with $B_i = B_{0i}$. By (2.2),

$$H_{0t} = C_{0w} + \sum_{i=1}^{r}\sum_{j=0}^{\infty} U'G_0^j U A_{0i}\vec{\varepsilon}_{0,t-i-j}, \tag{S.63}$$

where $C_{0w} = (I_m - \sum_{i=1}^{s} B_{0i})^{-1}W_0$, $U' = (I_m, 0_m, \cdots, 0_m)_{m\times ms}$, 0_m is an $m\times m$ matrix with all elements being 0. Since all the elements of B_{0i} are non-negative, it is not difficult to show that $U'G_0^jU \geq B_{01}^j$. Thus, we have

$$H_{0t} \geq C_{0w} + \sum_{i=1}^{r}\sum_{j=0}^{\infty} B_{01}^j A_{0i}\vec{\varepsilon}_{0,t-i-j} \tag{S.64}$$

$$\geq C_{0w} + \sum_{j=0}^{\infty} B_{01}^j A_{01}\vec{\varepsilon}_{0,t-j-1} \tag{S.65}$$

$$\geq C_{0w} + B_{01}^j A_{0i}\vec{\varepsilon}_{0,t-j-1}. \quad \text{for all j} \tag{S.66}$$

Note that $\mathbf{1'}\vec{\varepsilon}_{0,t-j-1} = \|\varepsilon_{0,t-j-1}\|^2$, where $\mathbf{1'} = (1,\ldots,1)$. Let $\tilde{\iota} \in (0, 1/2)$ such that $(\underline{b}m)^{\tilde{\iota}/2} > \rho$, where $\underline{b} = \min_{i,j} B_{01,ij}$, $B_{01,ij}$ is the (i,j)th entry of B_{01}, we have

$$\begin{aligned}
\frac{\xi_{0\varrho t-1}}{\sqrt{h_{0kt}}} \leq \frac{\xi_{0\varrho t-1}}{h_{0kt}^{\tilde{\iota}}} &\leq \frac{\xi_{0\varrho t-1}}{(e_k'C_{0w} + e_k'B_{01}^jA_{01}\vec{\varepsilon}_{0,t-j-1})^{\tilde{\iota}}} \\
&\leq C\left[1 + \sum_{j=0}^{\infty}\varrho^j \frac{\|\varepsilon_{0,t-j-1}\|}{\left[1 + e_k'B_{01}^j\mathbf{1}\mathbf{1'}\vec{\varepsilon}_{0,t-j-1}\right]^{\tilde{\iota}}}\right] \\
&= C\left[1 + \sum_{j=0}^{\infty}\varrho^j \frac{\|\varepsilon_{0,t-j-1}\|}{\left[1 + e_k'B_{01}^j\mathbf{1}\,\|\varepsilon_{0,t-j-1}\|^2\right]^{\tilde{\iota}}}\right] \\
&= C\left[1 + \sum_{j=0}^{\infty}\frac{\varrho^j}{(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}}}\|\varepsilon_{0,t-j-1}\|^{1-2\tilde{\iota}}\left(\frac{e_k'B_{01}^j\mathbf{1}\,\|\varepsilon_{0,t-j-1}\|^2}{1 + e_k'B_{01}^j\mathbf{1}\,\|\varepsilon_{0,t-j-1}\|^2}\right)^{\tilde{\iota}}\right] \\
&\leq C\left[1 + \sum_{j=0}^{\infty}\frac{\varrho^j}{(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}}}\|\varepsilon_{0,t-j-1}\|^{1-2\tilde{\iota}}(e_k'B_{01}^j\mathbf{1}\,\|\varepsilon_{0,t-j-1}\|^2)^{\tilde{\iota}/2}\right]
\end{aligned}$$

$$= C\left[1+\sum_{j=0}^{\infty}\frac{\varrho^j}{(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}}}\|\varepsilon_{0,t-j-1}\|^{1-2\tilde{\iota}}(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}/2}\|\varepsilon_{0,t-j-1}\|^{\tilde{\iota}}\right]$$

$$= C\left[1+\sum_{j=0}^{\infty}\frac{\varrho^j}{(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}/2}}\|\varepsilon_{0,t-j-1}\|^{1-\tilde{\iota}}\right]. \tag{S.67}$$

There exists $\varrho_1 \in (0,1)$,

$$\frac{\varrho^j}{(e_k'B_{01}^j\mathbf{1})^{\tilde{\iota}/2}} \le \frac{\varrho^j}{[(\underline{b}m)^j]^{\tilde{\iota}/2}} = \left(\frac{\varrho}{(\underline{b}m)^{\tilde{\iota}/2}}\right)^j \le \varrho_1^j. \tag{S.68}$$

By (S.67) and (S.68), using the same method as for (A.12) in Ling (2007), we have

$$\frac{\xi_{0\varrho t-1}}{\sqrt{h_{0kt}}} \le C\left[1+\sum_{j=0}^{\infty}\varrho_1^j\|\varepsilon_{0,t-j-1}\|^{1-\tilde{\iota}}\right] \le C\xi_{0\varrho_1 t-1}^{1-\tilde{\iota}},$$

for all $k = 1,\dots,m$. Thus, (ii) holds. This completes the proof. □

Lemma A.6 *If assumptions of Theorem 4.1 hold and* $\sqrt{n}\|\lambda_n - \lambda_0\| \le M$, *then it follows that*

(i) $\varepsilon_{nkt} = \varepsilon_{0kt} + o_p(1)\sqrt{h_{0kt}}$,

(ii) $h_{nkt} = h_{0kt} + o_p(1)h_{0kt}, \quad k = 1,\dots,m,$

where $\varepsilon_{nkt} = \varepsilon_{kt}(\varphi_n)$, $h_{nkt} = h_{kt}(\lambda_n)$, $o_p(1)$ *holds uniformly in* $t = 1,\dots,n$.

Proof. First, it is straightforward to show that

$$\varepsilon_{nkt} = \varepsilon_{0kt} + O(n^{-1/2})\xi_{\rho t-1}.$$

Furthermore, by Lemma A.5(i), we have a constant $\varrho \in (0,1)$ such that

$$\varepsilon_{nkt} = \varepsilon_{0kt} + O(n^{-1/2})\xi_{0\varrho t-1}. \tag{S.69}$$

By Theorem 2.6 in Zhang and Ling (2026), for any $\tilde{\iota}, \varrho \in (0,1)$, we have $E(\xi_{0\varrho t}^{1-\tilde{\iota}})^2 < \infty$. Thus,

$$\max_{1\le t\le n} \xi_{0\varrho t}^{1-\tilde{\iota}}/\sqrt{n} = o_p(1), \tag{S.70}$$

for any $\tilde{\iota}, \varrho \in (0,1)$. By (S.69)-(S.70), and Lemma A.5(ii), we can see (i) holds. Let G_n be defined as G in (2.5) with $\lambda = \lambda_n$, $\vec{\varepsilon}_{n,t-i-j} = \vec{\varepsilon}_{t-i-j}(\lambda_n)$, C_{nw} and A_{ni} be defined as C_{0w} and A_{0i} with $\lambda_0 = \lambda_n$. By (S.63),

$$\begin{aligned} h_{nkt} - h_{0kt} &= e_k'\left[C_{nw} - C_{0w} + \sum_{i=1}^{r}\sum_{j=1}^{\infty}(U'G_n^jUA_{ni} - U'G_0^jUA_{0i})\vec{\varepsilon}_{0,t-i-j}\right] \\ &\quad + e_k'\sum_{i=1}^{r}\sum_{j=1}^{\infty}U'G_n^jUA_{ni}\left[\vec{\varepsilon}_{n,t-i-j} - \vec{\varepsilon}_{0,t-i-j}\right] \\ &\equiv B_{1n} + B_{2n}, \end{aligned} \tag{S.71}$$

where U is defined as in (S.63). This is a constant $c > 0$ such that $G_0(1 - c/\sqrt{n}) \le G_n \le G_0(1 + c/\sqrt{n})$, where "$B \le C$" for matrices $B = (b_{ij})$ and $C = (c_{ij})$ means that $b_{ij} \le c_{ij}$ for all i and j. Thus, $U'G_n^jU - U'G_0^jU \le max\{|(1 - c/\sqrt{n})^j - 1|, |(1 + c/\sqrt{n})^j - 1|\}U'G_0^jU = O(j/\sqrt{n})(1 + c/\sqrt{n})^jU'G_0^jU$. Furthermore, since $e_k'(C_{nw} - C_{0w}) = O(n^{-1/2})$ and A_{ni} and A_{0i} are bounded, it follows that

$$\begin{aligned} B_{1n} &= e_k'[C_{nw} - C_{0w}] + e_k'\sum_{i=1}^{r}\sum_{j=1}^{\infty}U'G_n^jU(A_{ni} - A_{0i})\vec{\varepsilon}_{0,t-i-j} \\ &\quad + e_k'\sum_{i=1}^{r}\sum_{j=1}^{\infty}\left[U'G_n^jU - U'G_0^jU\right]A_{0i}\vec{\varepsilon}_{0,t-i-j} \end{aligned}$$

$$= O\left(\frac{1}{\sqrt{n}}\right) + e_k' O\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{r} \sum_{j=1}^{\infty} \left(1 + \frac{c}{\sqrt{n}}\right)^j U' G_0^j U \mathbf{1} \mathbf{1}' \vec{\varepsilon}_{0,t-i-j}$$
$$+ e_k' O\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{r} \sum_{j=1}^{\infty} j \left(1 + \frac{c}{\sqrt{n}}\right)^j U' G_0^j U \mathbf{1} \mathbf{1}' \vec{\varepsilon}_{0,t-i-j}$$
$$= O\left(\frac{1}{\sqrt{n}}\right) + e_k' O\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{r} \sum_{j=1}^{\infty} j \left(1 + \frac{c}{\sqrt{n}}\right)^j U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2 .$$

Using (S.63), as n is large enough, we have

$$\frac{1}{h_{0kt}} e_k' j \left(1 + \frac{c}{\sqrt{n}}\right)^j U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2$$
$$\leq j \left(1 + \frac{c}{\sqrt{n}}\right)^j \frac{e_k' U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2}{e_k' C_{0w} + e_k' U' G_0^j U A_{0i} \vec{\varepsilon}_{0,t-i-j}}$$
$$\leq Cj \left(1 + \frac{c}{\sqrt{n}}\right)^j \frac{e_k' U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2}{1 + e_k' U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2}$$
$$\leq Cj \left(1 + \frac{c}{\sqrt{n}}\right)^j (e_k' U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2)^{\iota_1}$$
$$\leq C\rho^j \left\|\varepsilon_{0,t-i-j}\right\|^{2\iota_1},$$

for some $\rho \in (0,1)$ and $\iota_1 \in (0, 1/2)$. Thus, as for (A.12) in Ling (2007), we can show that

$$B_{1n} = O\left(\frac{1}{\sqrt{n}}\right) + h_{0kt} O\left(\frac{1}{\sqrt{n}}\right) \sum_{i=1}^{r} \sum_{j=1}^{\infty} \rho^j \left\|\varepsilon_{0,t-i-j}\right\|^{2\iota_1}$$
$$= O\left(\frac{1}{\sqrt{n}}\right) \left[1 + h_{0kt} \sum_{i=1}^{r} \xi_{0\varrho t-i}^{2\iota_1}\right] = h_{0kt} o_p(1), \qquad \text{(S.72)}$$

by (S.70), where $\iota_1 \in (0, 1/2)$. By (S.63), $h_{0kt} \geq \underline{a} e_k' U' G_0^j U \mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\|^2$, where $\underline{a}$ is the minimum of entries of A_{0i}. Let $\overline{a}$ be the maximum of entries of A_{ni}, by (S.69), Lemma A.1(i) and Lemma A.5(i),

$$\begin{aligned}
B_{2n} &\leq \overline{a}e_k' \sum_{i=1}^{r}\sum_{j=1}^{\infty} U'G_n^j U\mathbf{1}\mathbf{1}'[\vec{\varepsilon}_{n,t-i-j} - \vec{\varepsilon}_{0,t-i-j}] \\
&\leq Ce_k' \left\{ \sum_{i=1}^{r}\sum_{j=1}^{\infty} U'G_n^j U\mathbf{1} \left\|\varepsilon_{n,t-i-j} - \varepsilon_{0,t-i-j}\right\| \left\|\varepsilon_{n,t-i-j} + \varepsilon_{0,t-i-j}\right\| \right\} \\
&\leq Ce_k' \left\{ \sum_{i=1}^{r}\sum_{j=1}^{\infty} U'G_n^j U\mathbf{1} \left[\frac{1}{\sqrt{n}}\xi_{0\varrho t-i-j-1}\right] \left[\left\|\varepsilon_{0,t-i-j}\right\| + \frac{1}{\sqrt{n}}\xi_{0\varrho t-i-j-1}\right] \right\} \\
&= Ce_k' \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{r}\sum_{j=1}^{\infty} U'G_n^j U\mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\| \xi_{0\varrho t-i-j-1} \right\} \\
&\quad + Ce_k' \left\{ \frac{1}{n} \sum_{i=1}^{r}\sum_{j=1}^{\infty} U'G_n^j U\mathbf{1} \xi^2_{0\varrho t-i-j-1} \right\} \\
&\leq C \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{r}\sum_{j=1}^{\infty} \left(1+\frac{c}{\sqrt{n}}\right)^j e_k' U'G_0^j U\mathbf{1} \left\|\varepsilon_{0,t-i-j}\right\| \xi_{0\varrho t-i-j-1} \right\} \\
&\quad + C \left\{ \frac{1}{n} \sum_{i=1}^{r}\sum_{j=1}^{\infty} \rho^j \xi^2_{0\varrho t-i-j-1} \right\} \\
&\leq C \left\{ \frac{\sqrt{h_{0kt}}}{\sqrt{n}} \sum_{i=1}^{r}\sum_{j=1}^{\infty} \left(1+\frac{c}{\sqrt{n}}\right)^j \left(e_k' U'G_0^j U\mathbf{1}\right)^{1/2} \xi_{0\varrho t-i-j-1} \right\} \\
&\quad + C \left\{ \frac{1}{n} \sum_{i=1}^{r}\sum_{j=1}^{\infty} \rho^j \xi^2_{0\varrho t-i-j-1} \right\} \\
&= O\left(\frac{1}{\sqrt{n}}\right) \sqrt{h_{0kt}} \sum_{i=1}^{r}\sum_{j=1}^{\infty} \rho^j \xi_{0\varrho t-i-j-1} + O\left(\frac{1}{n}\right) \sum_{i=1}^{r}\sum_{j=1}^{\infty} \left(\rho^j \xi_{0\varrho t-i-j-1}\right)^2,
\end{aligned}$$

for some $\rho, \varrho \in (0,1)$. Reordering $\sum_{i=1}^{r}\sum_{j=1}^{\infty}\sum_{l=1}^{\infty} \|\varepsilon_{0,t-i-j-l-1}\|$, we can show that $\exists \tilde{\varrho} \in (0,1)$, such that

$$B_{2n} = O(1)\left[\frac{\sqrt{h_{0kt}}}{\sqrt{n}}\xi_{0\tilde{\varrho}t-1} + \frac{1}{n}\xi^2_{0\tilde{\varrho}t-1}\right] = o_p(1)h_{0kt}, \qquad \text{(S.73)}$$

by (S.70) and Lemma A.5(ii), where $o_p(1)$ holds uniformly in $t = 1, \ldots, n$.

By (S.71)-(S.73), (ii) holds. This completes the proof. □

Lemma A.7 *If assumptions of Theorem 4.1 hold and* $\sqrt{n}\,\|\lambda_n - \lambda_0\| \leq M$*, then it follows that*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_n)}{\partial\lambda\partial\lambda'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda'} + o_p(1), \qquad \textit{for any fixed constant } M.$$

Proof. We first show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_n)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'} + o_p(1). \tag{S.74}$$

$\partial^2 l_t(\lambda)/\partial\tilde{\lambda}\partial\tilde{\lambda}'$ in (S.10) includes five terms, we only provide the proof of the following three equations, while other terms can be proved similarly.

$$\frac{1}{n}\sum_{t=1}^{n}R_t^{(2)}(\lambda_n) = \frac{1}{n}\sum_{t=1}^{n}R_t^{(2)}(\lambda_0) + o_p(1), \tag{S.75}$$

$$\frac{1}{n}\sum_{t=1}^{n}R_{1t}^{(3)}(\lambda_n) = \frac{1}{n}\sum_{t=1}^{n}R_{1t}^{(3)}(\lambda_0) + o_p(1), \tag{S.76}$$

$$\frac{1}{n}\sum_{t=1}^{n}R_{2t}^{(3)}(\lambda_n) = \frac{1}{n}\sum_{t=1}^{n}R_{2t}^{(3)}(\lambda_0) + o_p(1). \tag{S.77}$$

Denote $\tilde{\eta}_t$, η_t, Δ_t, $\tilde{\Delta}_t$, and D_t by $\tilde{\eta}_{nt}$, η_{nt}, Δ_{nt}, $\tilde{\Delta}_{nt}$, and D_{nt}, respectively, when $\lambda = \lambda_n$. By Lemma A.6,

$$\begin{aligned}|\eta_{nit} - \eta_{0it}| &\leq |\varepsilon_{nit} - \varepsilon_{0it}|\frac{1}{h_{nit}^{1/2}} + \left|h_{nit}^{1/2} - h_{0it}^{1/2}\right|\frac{|\varepsilon_{0it}|}{h_{nit}^{1/2}h_{0it}^{1/2}} \\ &= o_p(1) + o_p(1)|\eta_{0it}|, \\ \|\eta_{nt} - \eta_{0t}\| &= o_p(1) + o_p(1)\,\|\eta_{0t}\|\,, \end{aligned} \tag{S.78}$$

where $o_p(1)$ holds uniformly in all t, $i = 1, \ldots, m$. By (S.78),

$$
\begin{aligned}
&\left\|\tilde{\eta}_{nt}\Gamma_n^{-1}\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\Gamma_0^{-1}\tilde{\eta}_{0t}\right\| \\
&\quad \leq \left\|\tilde{\eta}_{nt}(\Gamma_n^{-1} - \Gamma_0^{-1})\tilde{\eta}_{nt}\right\| + \left\|(\tilde{\eta}_{nt} - \tilde{\eta}_{0t})\Gamma_0^{-1}(\tilde{\eta}_{nt} - \tilde{\eta}_{0t})\right\| \\
&\qquad + 2\left\|\tilde{\eta}_{0t}\Gamma_0^{-1}(\tilde{\eta}_{nt} - \tilde{\eta}_{0t})\right\| \\
&\quad = o_p(1) + o_p(1)\left\|\eta_{0t}\right\|^2, \qquad \text{(S.79)} \\
&\left\|\tilde{\Delta}_{nt}\tilde{\eta}_{nt} - \tilde{\Delta}_{0t}\tilde{\eta}_{0t}\right\| \\
&\quad \leq \left\|\tilde{\Delta}_{nt} - \tilde{\Delta}_{0t}\right\|\left\|\tilde{\eta}_{nt}\right\| + \left\|\tilde{\Delta}_{0t}\right\|\left\|\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\right\| \\
&\quad = \left\|\mathrm{diag}(e_1'\Gamma_n^{-1}\eta_{nt}, \ldots, e_m'\Gamma_n^{-1}\eta_{nt}) - \mathrm{diag}(e_1'\Gamma_0^{-1}\eta_{0t}, \ldots, e_m'\Gamma_0^{-1}\eta_{0t})\right\|\left\|\tilde{\eta}_{nt}\right\| \\
&\qquad + \left\|\mathrm{diag}(e_1'\Gamma_0^{-1}\eta_{0t}, \ldots, e_m'\Gamma_0^{-1}\eta_{0t})\right\|\left\|\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\right\| \\
&\quad \leq \left\|\mathrm{diag}(e_1'\Gamma_n^{-1}\eta_{nt}, \ldots, e_m'\Gamma_n^{-1}\eta_{nt}) - \mathrm{diag}(e_1'\Gamma_n^{-1}\eta_{0t}, \ldots, e_m'\Gamma_n^{-1}\eta_{0t})\right\|\left\|\tilde{\eta}_{nt}\right\| \\
&\qquad + \left\|\mathrm{diag}(e_1'\Gamma_n^{-1}\eta_{0t}, \ldots, e_m'\Gamma_n^{-1}\eta_{0t}) - \mathrm{diag}(e_1'\Gamma_0^{-1}\eta_{0t}, \ldots, e_m'\Gamma_0^{-1}\eta_{0t})\right\|\left\|\tilde{\eta}_{nt}\right\| \\
&\qquad + \left\|\mathrm{diag}(e_1'\Gamma_0^{-1}\eta_{0t}, \ldots, e_m'\Gamma_0^{-1}\eta_{0t})\right\|\left\|\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\right\| \\
&\quad = o_p(1) + o_p(1)\left\|\eta_{0t}\right\|^2. \qquad \text{(S.80)}
\end{aligned}
$$

By (S.79) and (S.80),

$$
\left\|\Delta_{nt} - \Delta_{0t}\right\| = o_p(1) + o_p(1)\left\|\eta_{0t}\right\|^2. \qquad \text{(S.81)}
$$

By Lemma A.2-A.3, and Lemma A.5, there exists a neighborhood Θ_0 of λ_0

such that

$$\sup_{\Theta_0}\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)D_t^{-2}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right\| \leq \sup_{\Theta_0}\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right\|^2 \leq C\xi_{0\varrho t-1}^{2-\tilde{\iota}}, \tag{S.82}$$

where C is a constant independent of $\tilde{\iota}$ and t. By (S.81)-(S.82), and Theorem 2.6 in Zhang and Ling (2026),

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}\Delta_{nt}D_{nt}^{-2}\frac{\partial H_{nt}}{\partial\tilde{\lambda}'} - \frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}\Delta_{0t}D_{nt}^{-2}\frac{\partial H_{nt}}{\partial\tilde{\lambda}'}\right]\right\| \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}(\Delta_{nt}-\Delta_{0t})D_{nt}^{-2}\frac{\partial H_{nt}}{\partial\tilde{\lambda}'}\right\| \\
&= \frac{1}{n}\sum_{t=1}^{n}\left[\xi_{0\varrho t-1}^{2-\tilde{\iota}}(o_p(1)+o_p(1)\left\|\eta_{0t}\right\|^2)\right] = o_p(1),
\end{aligned}$$

since $E\xi_{0\varrho t-1}^{2-\tilde{\iota}} < \infty$, $E\xi_{0\varrho t-1}^{2-\tilde{\iota}}\left\|\eta_{0t}\right\|^2 < \infty$, and $o_p(1)$ holds uniformly in $t = 1,\ldots,n$. By Lemma A.2-A.3, Lemma A.5, and Theorem 2.6 in Zhang and Ling (2026),

$$\begin{aligned}
&E\sup_{\Theta_0}\left\{\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right\|^2\left\|\Delta_{0t}\right\|\right\} \\
&\quad = O(1)E\sup_{\Theta_0}\left\{\left\|\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right\|^2\left\|\eta_{0t}\right\|^2\right\} < \infty.
\end{aligned}$$

Further, by proceeding equation, the dominated convergence theorem, and the ergodic theorem, we can show that (S.75) holds.

By (S.53) and (S.78), Lemmas A.2-A.3, Lemmas A.5, Theorem 2.6 in Zhang and Ling (2026),

$$\left\|\zeta_{nt}-\zeta_{0t}\right\| \leq \left\|\tilde{\eta}_{nt}\right\|\left\|\tilde{\eta}_{0t}\right\|\left\|\Gamma_n^{-1}-\Gamma_0^{-1}\right\|$$

$$+ 2 \left\|\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\right\| \left\|\Gamma_0^{-1}\eta_{0t}\right\| + \left\|\tilde{\eta}_{nt} - \tilde{\eta}_{0t}\right\|^2 \left\|\Gamma_0^{-1}\right\|$$

$$= o_p(1) + o_p(1) \left\|\eta_{0t}\right\|^2,$$

$$\sup_{\Theta} \left\| \frac{\partial}{\partial \tilde{\lambda}'} \text{vec}\left[\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial \tilde{\lambda}} D_t^{-2}(\lambda)\right]\right\| \leq C\xi_{0\varrho t-1}^{2-\tilde{\iota}}. \quad \text{(S.83)}$$

Thus,

$$\begin{aligned} &\left\|\frac{1}{n}\sum_{t=1}^{n}(\zeta_{nt}' \otimes I_N)\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}\right] - \frac{1}{n}\sum_{t=1}^{n}(\zeta_{0t}' \otimes I_N)\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}\right]\right\| \\ &\leq \frac{1}{n}\sum_{t=1}^{n}\left\|[(\zeta_{nt}' \otimes I_N) - (\zeta_{0t}' \otimes I_N)]\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_{nt}'}{\partial\tilde{\lambda}}D_{nt}^{-2}\right]\right\| \\ &= \frac{o_p(1)}{n}\sum_{t=1}^{n}\xi_{0\varrho t-1}^{2-\tilde{\iota}}[1 + \|\eta_{0t}\|^2] = o_p(1), \end{aligned} \quad \text{(S.84)}$$

where $o_p(1)$ holds uniformly in $t = 1, \ldots, n$.

Furthermore,

$$\begin{aligned} &E\sup_{\Theta_0}\left\{\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H_t'(\lambda)}{\partial\tilde{\lambda}}D_t^{-2}(\lambda)\right]\right\|\|\zeta_{0t} \otimes I_N\|\right\} \\ &= O(1)E[\xi_{0\varrho t-1}^{2-\tilde{\iota}}(1 + \|\eta_{0t}\|^2)] < \infty. \end{aligned} \quad \text{(S.85)}$$

By (S.84)-(S.85), the Dominated Convergence Theorem, and the ergodic theorem, (S.77) holds.

By (S.52),

$$\left(\varepsilon_t'(\varphi) \otimes I_N\right)\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}(D_t(\lambda)\Gamma D_t(\lambda))^{-1}\right]$$

$$
\begin{aligned}
&= [\varepsilon_t'(\varphi)D_t^{-1}(\lambda)\Gamma^{-1}D_t^{-1}(\lambda)\otimes I_N]\begin{pmatrix}\dfrac{\partial^2\varepsilon_{1t}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{\partial^2\varepsilon_{mt}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\end{pmatrix}\\
&\quad+\left[\varepsilon_t'(\varphi)\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}D_t^{-1}(\lambda)\Gamma^{-1}\right]\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&\quad+\left[\varepsilon_t'(\varphi)D_t^{-1}(\lambda)\Gamma^{-1}\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right]\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&= [\eta_t'(\lambda)\Gamma^{-1}D_t^{-1}(\lambda)\otimes I_N]\begin{pmatrix}\dfrac{\partial^2\varepsilon_{1t}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{\partial^2\varepsilon_{mt}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\end{pmatrix}\\
&\quad+\left[\eta_t'(\lambda)\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}D_t^{-1}(\lambda)\Gamma^{-1}\right]\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-2}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&\quad+\left[\eta_t'(\lambda)\Gamma^{-1}\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}\right]\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&= (\eta_t'(\lambda)\otimes I_N)(\Gamma^{-1}D_t^{-1}(\lambda)\otimes I_N)\begin{pmatrix}\dfrac{\partial^2\varepsilon_{1t}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\\ \vdots \\ \dfrac{\partial^2\varepsilon_{mt}(\varphi)}{\partial\tilde{\lambda}\partial\tilde{\lambda}'}\end{pmatrix}\\
&\quad+(\eta_t'(\lambda)\otimes I_N)(I_m\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}}D_t^{-1}(\lambda)\Gamma^{-1})\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-2}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&\quad+(\eta_t'(\lambda)\otimes I_N)(\Gamma^{-1}\otimes\frac{\partial\varepsilon_t'(\varphi)}{\partial\tilde{\lambda}})\left[-\frac{1}{2}\mathrm{diag}(e_1,\ldots,e_m)D_t^{-3}(\lambda)\frac{\partial H_t(\lambda)}{\partial\tilde{\lambda}'}\right]\\
&\equiv (\eta_t'(\lambda)\otimes I_N)A_t(\lambda), \qquad\qquad (S.86)
\end{aligned}
$$

where

$$A_t(\lambda) = (\Gamma^{-1} D_t^{-1}(\lambda) \otimes I_N) \begin{pmatrix} \dfrac{\partial^2 \varepsilon_{1t}(\varphi)}{\partial \tilde{\lambda} \partial \tilde{\lambda}'} \\ \vdots \\ \dfrac{\partial^2 \varepsilon_{mt}(\varphi)}{\partial \tilde{\lambda} \partial \tilde{\lambda}'} \end{pmatrix}$$
$$+ (I_m \otimes \frac{\partial \varepsilon_t'(\varphi)}{\partial \tilde{\lambda}} D_t^{-1}(\lambda)\Gamma^{-1}) \left[-\frac{1}{2} \text{diag}(e_1, \ldots, e_m) D_t^{-2}(\lambda) \frac{\partial H_t(\lambda)}{\partial \tilde{\lambda}'} \right]$$
$$+ (\Gamma^{-1} \otimes \frac{\partial \varepsilon_t'(\varphi)}{\partial \tilde{\lambda}}) \left[-\frac{1}{2} \text{diag}(e_1, \ldots, e_m) D_t^{-3}(\lambda) \frac{\partial H_t(\lambda)}{\partial \tilde{\lambda}'} \right].$$

So (S.76) is equivalently to

$$\frac{1}{n} \sum_{t=1}^{n} (\eta_{nt}' \otimes I_N) A_{nt} = \frac{1}{n} \sum_{t=1}^{n} (\eta_{0t}' \otimes I_N) A_{0t} + o_p(1). \tag{S.87}$$

By (S.78), Lemmas A.1-A.3, A.5, and Theorem 2.6 in Zhang and Ling (2026),

$$\sup_{\Theta_0} \|A_t(\lambda)\| \leq C \xi_{0\varrho t-1}^{2-\tilde{\iota}}, \tag{S.88}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} [(\eta_{nt}' \otimes I_N) A_{nt} - (\eta_{0t}' \otimes I_N) A_{nt}] \right\|$$
$$= \frac{1}{n} \sum_{t=1}^{n} \|[(\eta_{nt}' - \eta_{0t}') \otimes I_N] A_{nt}\|$$
$$= \frac{1}{n} \sum_{t=1}^{n} [(o_p(1) + o_p(1) \|\eta_{0t}\|) \xi_{0\varrho t-1}^{2-\tilde{\iota}}] = o_p(1). \tag{S.89}$$

Furthermore,

$$E \sup_{\Theta_0} \{ \|A_t(\lambda)\| \, \|\eta_{0t} \otimes I_N\| \} \leq O(1) E[\xi_{0\varrho t-1}^{2-\tilde{\iota}} \|\eta_{0t}\|] < \infty. \tag{S.90}$$

By the (S.89)-(S.90), the Dominated Convergence Theorem, and the ergodic theorem, (S.87) holds, and then (S.76) holds.

Similarly, we can show

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_n)}{\partial\varphi\partial\sigma'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\varphi\partial\sigma'} + o_p(1),$$
$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_n)}{\partial\delta\partial\sigma'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\delta\partial\sigma'} + o_p(1),$$
$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_n)}{\partial\sigma\partial\sigma'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\sigma\partial\sigma'} + o_p(1).$$

This completes the proof. □

Proof of Thereom 4.1

Proof. We first show that

$$E\left\|\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda}\right\| < \infty \qquad \text{and} \qquad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda'} = \Sigma + o_p(1). \tag{S.91}$$

By (S.10)-(S.14), (S.83), (S.86), (S.88), Lemmas A.1-A.3, Lemma A.5, and Theorem 2.6 in Zhang and Ling (2026), it follows that

$$E\left\|\frac{\partial\varepsilon_{0t}'}{\partial\tilde{\lambda}}(D_{0t}\Gamma_0 D_{0t})^{-1}\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\| \le CE\left\|\frac{\partial\varepsilon_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-1}\right\|^2 < \infty,$$
$$E\left\|\frac{\partial H_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-2}\Delta_{0t}D_{0t}^{-2}\frac{\partial H_{0t}}{\partial\tilde{\lambda}'}\right\| \le CE\left\|\frac{\partial H_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-2}\right\|^2 E\left\|\Delta_{0t}\right\| < \infty,$$
$$E\left\|\frac{\partial H_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-2}[\tilde{\eta}_{0t}\Gamma_0^{-1}D_{0t}^{-1} + \tilde{\Delta}_{0t}D_{0t}^{-1}]\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\|$$
$$\le E\left\|\frac{\partial H_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-2}\tilde{\eta}_{0t}\Gamma_0^{-1}D_{0t}^{-1}\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\| + E\left\|\frac{\partial H_{0t}'}{\partial\tilde{\lambda}}D_{0t}^{-2}\tilde{\Delta}_{0t}D_{0t}^{-1}\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\|$$

$$= O(1)E\left\|\frac{\partial H'_{0t}}{\partial\tilde{\lambda}}D_{0t}^{-2}\right\|\left\|D_{0t}^{-1}\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\|E\left\|\tilde{\eta}_{0t}\right\|$$

$$+ O(1)E\left\|\frac{\partial H'_{0t}}{\partial\tilde{\lambda}}D_{0t}^{-2}\right\|\left\|D_{0t}^{-1}\frac{\partial\varepsilon_{0t}}{\partial\tilde{\lambda}'}\right\|E\left\|\tilde{\Delta}_{0t}\right\| < \infty,$$

$$E\left\|(\zeta_{0t}\otimes I_N)\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H'_{0t}}{\partial\tilde{\lambda}}D_{0t}^{-2}\right]\right\|$$

$$\leq E\left\|\zeta_{0t}\otimes I_N\right\|E\left\|\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{1}{2}\frac{\partial H'_{0t}}{\partial\tilde{\lambda}}D_{0t}^{-2}\right]\right\| < \infty,$$

$$E\left\|(\varepsilon'_{0t}\otimes I_N)\frac{\partial}{\partial\tilde{\lambda}'}\text{vec}\left[\frac{\partial\varepsilon'_{0t}(\lambda)}{\partial\tilde{\lambda}}(D_{0t}\Gamma_0 D_{0t})^{-1}\right]\right\|$$

$$\leq E\left\|\eta'_{0t}\otimes I_N\right\|E\left\|A_{0t}\right\| < \infty.$$

Similarly, we can show that the other terms in $E\left[\partial^2 l_t(\lambda_0)/\partial\lambda\partial\lambda'\right]$ are finite. Thus, the first part in (S.91) holds. The second part of (S.91) holds by the ergodic theorem. By Taylor's expansion, we have

$$\sum_{t=1}^{n}\frac{\partial l_t^{\epsilon}(\hat{\lambda}_{sn})}{\partial\lambda} = \sum_{t=1}^{n}\frac{\partial l_t^{\epsilon}(\lambda_0)}{\partial\lambda} + \sum_{t=1}^{n}\frac{\partial^2 l_t^{\epsilon}(\lambda_n^*)}{\partial\lambda\partial\lambda'}(\hat{\lambda}_{sn}-\lambda_0), \tag{S.92}$$

where λ_n^* lies between λ_0 and $\hat{\lambda}_{sn}$. By (S.91), Lemma A.4(iii), and Lemma A.7, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t^{\epsilon}(\hat{\lambda}_{sn})}{\partial\lambda\partial\lambda'} = \Sigma + o_p(1) \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t^{\epsilon}(\lambda_n^*)}{\partial\lambda\partial\lambda'} = \Sigma + o_p(1). \tag{S.93}$$

By Lemma A.4(ii), we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t^{\epsilon}(\lambda_0)}{\partial\lambda} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\lambda_0)}{\partial\lambda} + o_p(1). \tag{S.94}$$

As in Ling and McAleer (2003) and Francq and Zakoïan (2004), we can show that $-\Sigma > 0$ and $\Omega > 0$. Furthermore, since $\sqrt{n}(\hat{\lambda}_{sn}-\lambda_0) = O_p(1)$, by (4.1), (S.92)-(S.94), we have

$$\hat{\lambda}_n = \hat{\lambda}_{sn} - [\Sigma + o_p(1)]^{-1} \left\{ \frac{1}{n} \sum_{t=1}^{n} \frac{\partial l_t(\lambda_0)}{\partial \lambda} + [\Sigma + o_p(1)](\hat{\lambda}_{sn} - \lambda_0) \right\} + o_p\left(\frac{1}{\sqrt{n}}\right)$$
$$= \lambda_0 - \frac{\Sigma^{-1}}{n} \sum_{t=1}^{n} \frac{\partial l_t(\lambda_0)}{\partial \lambda} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Finally, by the central limiting theorem, we can show that the conclusion holds. This completes the proof. □

Proof of Theorem 5.1

Proof. If the model is correct, by (3.2), (5.2), Theorem 3.1 in Ling and McAleer (2003), the dominated convergence theorem, and similar argument to Lemma A.4,

$$\bar{\varsigma}_n = E\varsigma_{0\omega t} + o_p(1) = o_p(1),$$
$$\frac{1}{n} \sum_{t=1}^{n} [\varsigma_{\omega t}^{\epsilon}(\hat{\lambda}_{sn}) - \bar{\varsigma}_n]^2 = E\varsigma_{0\omega t}^2 + o_p(1) = mkE\omega_t^2 + o_p(1),$$

and hence

$$\hat{R}_{sn,l} = \frac{\sum_{t=l+1}^{n} \varsigma_{\omega t}(\hat{\lambda}_{sn})\varsigma_{\omega t-l}(\hat{\lambda}_{sn})/n}{mkE\omega_t^2} + o_p(1). \tag{S.95}$$

We only need to consider the asymptotic distribution of

$$\hat{K}_{sn,l} = \frac{1}{n} \sum_{t=l+1}^{n} \varsigma_{\omega t}(\hat{\lambda}_{sn})\varsigma_{\omega t-l}(\hat{\lambda}_{sn}). \tag{S.96}$$

Let $K = (K_1, K_2, \ldots, K_M)'$ and $\hat{K}_{sn} = (\hat{K}_{sn,1}, \hat{K}_{sn,2}, \ldots, \hat{K}_{sn,M})'$, where

$\hat{K}_{sn,l}$ is K_l when $\hat{\lambda}_{sn}$ is replaced by λ_0. By Taylor's expansion, we have

$$\hat{K}_{sn} = K + \frac{\partial K}{\partial \lambda'}(\hat{\lambda}_{sn} - \lambda_0) + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{S.97}$$

where $\partial K/\partial \lambda' = (\partial K_1/\partial \lambda, \partial K_2/\partial \lambda, \ldots, \partial K_M/\partial \lambda)'$, and

$$\begin{aligned}
\frac{\partial K_l}{\partial \lambda} = & \frac{1}{n}\sum_{t=l+1}^{n} \varsigma_{0\omega,t}\left[\omega_{t-l}\frac{2\partial \varepsilon'_{0,t-l}}{\partial \lambda}(D_{0,t-l}\Gamma_0 D_{0,t-l})^{-1}\varepsilon_{0,t-l}\right] \\
& - \frac{1}{n}\sum_{t=l+1}^{n} \varsigma_{0\omega,t}\Bigg\{\omega_{t-l}\frac{\partial \mathrm{vec}(D_{0,t-l}\Gamma_0 D_{0,t-l})'}{\partial \lambda} \\
& \qquad \times \mathrm{vec}[(D_{0,t-l}\Gamma_0 D_{0,t-l})^{-1}\varepsilon_{0,t-l}\varepsilon'_{0,t-l}(D_{0,t-l}\Gamma_0 D_{0,t-l})^{-1}]\Bigg\} \\
& + \frac{1}{n}\sum_{t=l+1}^{n} \varsigma_{0\omega,t-l}\left[\omega_t\frac{2\partial \varepsilon'_{0t}}{\partial \lambda}(D_{0t}\Gamma_0 D_{0t})^{-1}\varepsilon_{0t}\right] \\
& - \frac{1}{n}\sum_{t=l+1}^{n} \varsigma_{0\omega,t-l}\left\{\omega_t\frac{\partial \mathrm{vec}(D_{0t}\Gamma_0 D_{0t})'}{\partial \lambda}\mathrm{vec}[(D_{0t}\Gamma_0 D_{0t})^{-1}\varepsilon_{0t}\varepsilon'_{0t}(D_{0t}\Gamma_0 D_{0t})^{-1}]\right\}.
\end{aligned} \tag{S.98}$$

By the ergodic theorem, it is easy to obtain

$$\frac{\partial K_l}{\partial \lambda} \xrightarrow{a.s.} -X_l, \quad \text{as} \quad n \to \infty. \tag{S.99}$$

By (S.99), (S.97) can be written as

$$\hat{K}_{sn} = K - X(\hat{\lambda}_{sn} - \lambda_0) + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.100}$$

According to the proof of Theorem 3.1 (ii),

$$\hat{\lambda}_{sn} - \lambda_0 = -\left[\Sigma_0 + o_p(1)\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\omega_t\frac{\partial l_t(\lambda_0)}{\partial \lambda}\right]. \tag{S.101}$$

By (S.95), (S.100)-(S.101),

$$
\begin{aligned}
\Big[mkE\omega_t^2\Big][\sqrt{n}\hat{R}_{sn}] &= \sqrt{n}\hat{K}_{sn} + o_p(1) \\
&= \sqrt{n}K - X[\sqrt{n}(\hat{\lambda}_{sn} - \lambda_0)] + o_p(1) \\
&= \sqrt{n}K - X\left[-\Sigma_0^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\omega_t\frac{\partial l_t(\lambda_0)}{\partial\lambda}\right)\right] + o_p(1) \\
&= \sqrt{n}K + X\Sigma_0^{-1}\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n}\omega_t\frac{\partial l_t(\lambda_0)}{\partial\lambda}\right) + o_p(1) \\
&= VZ_n^{\diamond} + o_p(1),
\end{aligned}
$$

where $Z_n^{\diamond} = \sqrt{n}\left\{K, \frac{1}{n}\sum_{t=1}^{n}\omega_t\frac{\partial l_t(\lambda_0)}{\partial\lambda'}\right\}'$.

Finally, the conclusion holds by the central limit theorem for martingale difference sequence. □

Bibliography

Amemiya, T. (1985). Advanced Econometrics. *Harvard university press.*

Francq, C. and Zakoian, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli ***10****, 605–637.*

Francq, C. and Zakoian, J. M. (2019). GARCH models: structure, statistical inference and financial applications. *2nd Edition. John wiley & sons.*

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estima-

tors for ARMA-GARCH/IGARCH models. Journal of Econometrics ***140***, *849–873.*

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory ***19***, *280–310.*

Zhang, R. and Ling, S. (2026). Inference for vector AR models with heavy-tailed vector GARCH noises. Science China Mathematics, *1–16.*