Supplementary material for
"Hybrid quantile regression estimation for time series models with conditional heteroscedasticity"

Yao Zheng, Qianqian Zhu, Guodong Li and Zhijie Xiao

University of Hong Kong, Shanghai University of Finance and Economics and Boston College

Abstract

This supplementary material contains theoretical comparison with the unweighted and FHS methods, additional simulation results on the finite-sample performance of the proposed estimation method in comparison with existing and unweighted methods and on the random weights in the bootstrapping procedure, as well as additional results for the empirical analysis. It also provides technical details for Lemma A.1, Theorems 1–4, Corollaries 1–3 and Equation (2.6).

1 Theoretical comparison with the unweighted and FHS methods

In this section, we compare the asymptotic efficiency of the proposed estimator $\widehat{\theta}_{\tau n}$ with that of the unweighted estimator $\breve{\theta}_{\tau n}$ and that of the FHS estimator $\widetilde{\theta}_{\tau n}$.

To compare the asymptotic efficiency of the proposed estimator $\widehat{\theta}_{\tau n}$ and its unweighted counterpart $\breve{\theta}_{\tau n}$, we calculate the asymptotic relative efficiency (ARE) of $\widehat{\theta}_{\tau n}$ to $\breve{\theta}_{\tau n}$, defined as $\text{ARE}(\widehat{\theta}_{\tau n}, \breve{\theta}_{\tau n}) = (|\Sigma_2|/|\Sigma_1|)^{1/(p+q+1)}$, where Σ_1 and Σ_2 are the asymptotic covariance matrices of $\widehat{\theta}_{\tau n}$ and $\breve{\theta}_{\tau n}$, respectively, and $|\cdot|$ is the determinant of a matrix; see Serfling (1980). As Σ_1 and Σ_2 both depend on the GARCH parameters, the innovation distribution and the quantile level in a very complicated way, we will consider the AREs for specific settings.

Table S.1: $\text{ARE}(\widehat{\theta}_{\tau n}, \breve{\theta}_{\tau n})$ for the GARCH(1, 1) model with $\alpha_0 = 0.1$ and different values for (α_1, β_1), where the innovations $\{\eta_t\}$ follow the standard normal and Student's $t_{4.1}$ and t_5 distributions, and $\tau = 0.01$, 0.05, 0.1 and 0.15, based on a generated sequence of $n = 10,000$.

τ	β_1	0.15			0.45			0.80
	α_1	0.15	0.45	0.80	0.15	0.45	0.80	0.15
0.01	$t_{4.1}$	1.39	3.31	5.14	1.28	4.19	35.11	1.79
	t_5	1.23	1.69	9.77	1.39	3.35	71.09	11.02
	Normal	1.06	1.45	9.75	1.07	7.17	57.92	1.53
0.05	$t_{4.1}$	1.32	8.51	42.83	2.58	4.74	21.07	1.89
	t_5	1.79	2.22	9.15	1.19	10.11	23.54	1.85
	Normal	1.04	1.63	4.42	1.09	3.46	35.54	1.43
0.10	$t_{4.1}$	2.55	4.06	5.70	1.29	4.11	28.85	3.46
	t_5	1.15	2.36	17.13	1.48	12.04	80.63	1.77
	Normal	1.06	1.43	7.33	1.08	4.61	105.47	1.42
0.15	$t_{4.1}$	2.25	2.23	6.43	1.41	2.28	21.16	1.45
	t_5	1.66	6.17	36.80	2.00	3.85	69.58	2.40
	Normal	1.05	1.72	9.69	1.07	2.19	92.47	1.35

We generate a sequence with $n = 10,000$ from the GARCH(1, 1) model,

$$x_t = \sqrt{h_t}\eta_t, \quad h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}, \tag{S.1}$$

where $\alpha_0 = 0.1$, and the innovations $\{\eta_t\}$ follow the standard normal and standardized Student's t_5 and $t_{4.1}$ distributions with unit variance. To calculate Σ_1 and Σ_2, we substitute the matrices involved in them by corresponding sample-average estimates and use theoretical values for b_τ, $f(b_\tau)$, κ_1 and κ_2, since both $f(\cdot)$ and τ are known. We consider different values for (α_1, β_1), and the results are given in Table S.1. It can be seen that $\text{ARE}(\widehat{\theta}_{\tau n}, \breve{\theta}_{\tau n}) > 1$ for all cases considered; i.e., the proposed weighted estimator is asymptotically more efficient than the unweighted estimator.

Likewise, we can compute the ARE of the proposed estimator $\widehat{\theta}_{\tau n}$ to the FHS estimator $\widetilde{\theta}_{\tau n}$, and the results are reported in Table S.2. It can be seen that the FHS estimator $\widetilde{\theta}_{\tau n}$ is asymptotically more efficient, i.e., $\text{ARE}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_{\tau n}) < 1$, when $\{\eta_t\}$ follow the Student's t_5 distribution, while the proposed estimator $\widehat{\theta}_{\tau n}$ can be asymptotically more efficient, i.e., $\text{ARE}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_{\tau n}) > 1$, when $\{\eta_t\}$ become more heavy-tailed. This can be explained in part by the efficiency gain of the quantile regression estimation and the

Table S.2: ARE($\widehat{\theta}_{\tau n}, \widetilde{\theta}_{\tau n}$) for the GARCH(1,1) model with $\alpha_0 = 0.1$ and different values for (α_1, β_1), where the innovations $\{\eta_t\}$ follow the Student's $t_{4.1}$, $t_{4.5}$ and t_5 distributions, and $\tau = 0.01$, 0.05, 0.1 and 0.15, based on a generated sequence of $n = 10,000$.

τ	β_1	0.15			0.45			0.80
	α_1	0.15	0.45	0.80	0.15	0.45	0.80	0.15
0.01	$t_{4.1}$	1.32	1.26	1.24	0.85	0.80	0.77	0.30
	$t_{4.5}$	0.54	0.52	0.52	0.38	0.37	0.35	0.14
	t_5	0.40	0.39	0.38	0.30	0.27	0.26	0.11
0.05	$t_{4.1}$	2.47	2.35	2.27	1.45	1.32	1.27	0.41
	$t_{4.5}$	1.02	0.99	0.97	0.71	0.67	0.62	0.24
	t_5	0.75	0.72	0.71	0.53	0.50	0.48	0.20
0.10	$t_{4.1}$	2.75	2.61	2.53	1.58	1.40	1.38	0.48
	$t_{4.5}$	1.14	1.10	1.08	0.76	0.71	0.69	0.26
	t_5	0.81	0.80	0.77	0.59	0.54	0.52	0.21
0.15	$t_{4.1}$	2.62	2.49	2.41	1.47	1.37	1.33	0.45
	$t_{4.5}$	1.07	1.05	1.02	0.73	0.68	0.66	0.25
	t_5	0.77	0.74	0.73	0.55	0.50	0.48	0.19

efficiency loss of the Gaussian QMLE, as the data become more heavy-tailed. In addition, for a given parameter vector and innovation distribution, it can be observed that the ARE is generally the largest when $\tau = 0.05$ and 0.1.

2 Additional simulation results

2.1 Finite-sample comparison of conditional quantile estimation performance with existing methods

In this subsection, we focus on three data generating processes as follows.

- Model 1 (Global GARCH process with larger volatility):
  i.e., model (S.1) with $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.8, 0.15)$;

- Model 2 (Global GARCH process with more persistent effect of shocks):
  i.e., model (S.1) with $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.15, 0.8)$;

- Model 3 (Quantile process which contains the GARCH process as a special case):

$$x_t = \Phi^{-1}(U_t)\sqrt{\frac{\alpha_0}{1-\beta_1} + [\alpha_{11}I(U_t \geqslant 0.5) + \alpha_{12}I(U_t < 0.5)]\sum_{j=1}^{\infty}\beta_1^{j-1}x_{t-j}^2},$$

  where $\{U_t\}$ are *i.i.d.* uniform over $(0,1)$, $\Phi(\cdot)$ is the distribution function of the standard normal distribution, and $(\alpha_0, \alpha_{11}, \alpha_{12}, \beta_1) = (0.1, 0.8, 0.15, 0.15)$.

Note that for model (S.1) with standard normal innovations $\{\eta_t\}$, the conditional quantile function is

$$Q_\tau(x_t|\mathcal{F}_{t-1}) = \Phi^{-1}(\tau)\sqrt{\frac{\alpha_0}{1-\beta_1} + \alpha_1\sum_{j=1}^{\infty}\beta_1^{j-1}x_{t-j}^2},$$

which is a special case of Model 3 with $\alpha_{11} = \alpha_{12} = \alpha_1$. However, when $\alpha_{11} \neq \alpha_{12}$, Model 3 allows not only the scale $Q_{\tau,\eta} = \Phi^{-1}(\tau)$ of the conditional quantile $Q_\tau(x_t|\mathcal{F}_{t-1})$ to change with τ, but also its shape, since

$$Q_\tau(x_t|\mathcal{F}_{t-1}) = \begin{cases} \Phi^{-1}(\tau)\sqrt{\frac{\alpha_0}{1-\beta_1} + \alpha_{11}\sum_{j=1}^{\infty}\beta_1^{j-1}x_{t-j}^2}, & \text{if } \tau \geqslant 0.5 \\ \Phi^{-1}(\tau)\sqrt{\frac{\alpha_0}{1-\beta_1} + \alpha_{12}\sum_{j=1}^{\infty}\beta_1^{j-1}x_{t-j}^2}, & \text{if } \tau < 0.5 \end{cases}.$$

We first consider the global GARCH processes, i.e., Models 1 and 2, with the innovations $\{\eta_t\}$ following the standard normal or standardized Student's t_5 distribution. We estimate the conditional quantiles at $\tau = 0.05$ using six estimation methods: the proposed hybrid method, the FHS method, and the four other methods discussed in Section 6 of the paper. We call the estimates of $Q_\tau(x_t|\mathcal{F}_{t-1})$ for $1 \leqslant t \leqslant n$ the in-sample forecasts, and that of $Q_\tau(x_{n+1}|\mathcal{F}_n)$ the out-of-sample forecast. Three sample sizes, $n = 200$, 500 and 1000, are considered, and 1000 replications are generated for each sample size. For each setting, we compute the bias and mean squared error (MSE) of the estimates by averaging individual values over all time points and all samples. The results for Models 1 and 2 are reported in Tables S.3 and S.4, respectively.

Before comparing the specific methods in details, we list several general observations from Tables S.3 and S.4: (1) a smaller in-sample bias or MSE is usually associated with a smaller out-of-sample bias or MSE; (2) for all methods except RiskM, the absolute value of the in-sample bias and in-sample MSE generally decrease as n increases, while the out-of-sample performance is less stable; (3) the RiskM method performs significantly poorer than the other methods in terms of both the bias and MSE in most cases. Therefore, in the following comparison, we will leave the RiskM method aside.

Table S.3: Bias (×10) and MSE for in-sample and out-of-sample conditional quantile estimates obtained by six methods at $\tau = 0.05$ for Model 1 with normal or Student's t_5-distributed innovations.

		Normal distribution				Student's t_5 distribution			
		Bias		MSE		Bias		MSE	
n		In	Out	In	Out	In	Out	In	Out
200	Hybrid	-0.028	-0.020	0.121	0.088	-0.231	-0.094	0.194	0.175
	FHS	-0.271	-0.275	0.075	0.057	-0.509	-0.474	0.139	0.107
	XK_1	0.293	0.130	0.390	0.275	0.131	0.115	0.472	0.417
	XK_2	0.300	0.134	0.368	0.319	0.137	0.066	0.475	0.638
	CAViaR	0.165	0.060	0.162	0.147	-0.060	-0.035	0.291	0.270
	RiskM	-1.266	-1.572	1.633	1.261	-1.491	-1.818	1.338	1.324
500	Hybrid	-0.017	0.004	0.064	0.046	-0.079	-0.070	0.092	0.049
	FHS	-0.108	-0.099	0.041	0.029	-0.198	-0.167	0.073	0.028
	XK_1	0.201	0.205	0.354	0.139	0.132	0.077	0.430	0.134
	XK_2	0.205	0.219	0.358	0.137	0.148	0.060	0.447	0.134
	CAViaR	0.059	0.043	0.128	0.066	0.009	0.014	0.273	0.070
	RiskM	-1.591	-1.585	2.282	1.467	-1.615	-1.745	1.603	1.162
1000	Hybrid	-0.001	-0.007	0.028	0.023	-0.040	-0.047	0.048	0.032
	FHS	-0.045	-0.054	0.013	0.010	-0.105	-0.231	0.032	0.061
	XK_1	0.153	0.090	0.279	0.173	0.127	0.557	0.414	12.911
	XK_2	0.152	0.110	0.271	0.147	0.130	0.500	0.422	10.190
	CAViaR	0.037	0.026	0.075	0.039	0.001	0.057	0.198	0.205
	RiskM	-1.566	-1.700	1.951	1.472	-1.637	-1.492	1.931	2.897

We first compare the bias for different methods. For both models, the proposed hybrid method has the smallest bias when $\{\eta_t\}$ are normal, while the CAViaR method has the smallest bias when $\{\eta_t\}$ are Student's t_5-distributed. This may be explained by the greater efficiency of the Gaussian QMLE, which is employed in Step E1, for normal innovations than the Student's t_5-distributed innovations. It is also clear that the FHS method has much larger biases than the proposed method and the CAViaR method for Model 1, and it has much larger biases than all quantile regression based methods for Model 2. It is worth pointing out that the generally smaller biases of the quantile regression based methods reflect their greater flexibility in capturing the specific conditional quantile structure.

Table S.4: Bias (×10) and MSE for in-sample and out-of-sample conditional quantile estimates obtained by six methods at $\tau = 0.05$ for Model 2 with normal or Student's t_5-distributed innovations.

		Normal distribution				Student's t_5 distribution			
		Bias		MSE		Bias		MSE	
n		In	Out	In	Out	In	Out	In	Out
200	Hybrid	-0.193	-0.268	0.193	0.207	-0.593	-0.726	0.401	0.461
	FHS	-0.685	-0.755	0.114	0.120	-1.194	-1.352	0.262	0.278
	XK_1	-0.103	-0.112	0.392	0.471	-0.417	-0.533	0.741	0.866
	XK_2	-0.075	-0.012	0.350	0.422	-0.333	-0.360	0.660	0.835
	CAViaR	0.129	0.218	0.157	0.194	-0.143	-0.079	0.317	0.365
	RiskM	0.466	-0.061	0.150	0.142	-0.460	-1.017	0.270	0.272
500	Hybrid	-0.027	0.034	0.078	0.082	-0.166	-0.105	0.145	0.166
	FHS	-0.253	-0.218	0.045	0.048	-0.442	-0.458	0.090	0.086
	XK_1	-0.061	0.071	0.231	0.266	-0.166	-0.102	0.435	0.561
	XK_2	-0.017	0.085	0.173	0.191	-0.129	-0.076	0.342	0.613
	CAViaR	0.099	0.181	0.069	0.078	0.006	0.110	0.131	0.156
	RiskM	0.249	0.167	0.132	0.128	-0.580	-0.581	0.236	0.207
1000	Hybrid	0.002	-0.006	0.038	0.041	-0.084	-0.172	0.077	0.132
	FHS	-0.092	-0.097	0.021	0.021	-0.216	-0.276	0.048	0.094
	XK_1	-0.068	-0.020	0.146	0.155	-0.156	-0.348	0.361	1.334
	XK_2	-0.020	0.010	0.097	0.103	-0.100	-0.298	0.259	1.254
	CAViaR	0.066	0.073	0.034	0.038	-0.001	-0.001	0.092	0.085
	RiskM	0.175	0.090	0.129	0.128	-0.627	-0.597	0.247	0.287

It is also noteworthy that the XK methods perform poorly for Model 1, but have fairly small bias for Model 2. This is caused by the sieve approximation $h_t = \gamma_0 + \sum_{j=1}^{m} \gamma_j x_{t-j}^2$ in the XK methods, where an unnecessarily large order m can introduce too much noise. Notice that a larger n needs a larger m, and smaller α_1 and β_1 favor smaller m. As the magnitude of β_1 has a greater impact on the choice of m than α_1, the problem of choosing an excessively large m is more severe in Model 1.

For the MSE, the FHS method is the best method in most cases, which not surprising since when the true model is the GARCH model, the FHS method is expected to be generally more efficient than any quantile regression based method. The second best method in terms of the MSE is the proposed hybrid method for Model 1, and is the

CAViaR method for Model 2. This is also as expected for the reason that, compared with CAViaR, the proposed hybrid method relies on an initial estimation that reduces efficiency, but uses weights to improve efficiency. As a result, the efficiency gain from the weights will be more pronounced when the conditional variances $\{h_t\}$ have larger variations, namely the case of Model 1. In addition, it is noteworthy that the proposed procedure takes much less computation time than CAViaR. For instance, for our 1000 replications of Model 1 with normal innovations and n =1000, CAViaR takes 15.6 minutes, but the proposed procedure takes only 2.8 minutes.

Finally, we consider Model 3, which is a quantile process and possesses different shapes for $\tau > 0.5$ and $\tau < 0.5$. Thus, the conditional quantile structure is misspecified if a GARCH$(1,1)$ model is assumed. We estimate the conditional quantiles of the 1000 replications generated from Model 3 using exactly the same estimation methods as for

Table S.5: Bias ($\times 10$) and MSE ($\times 10$) for in-sample and out-of-sample conditional quantile estimates obtained by six methods at $\tau = 0.05$ for Model 3.

		Bias		MSE	
n		In	Out	In	Out
200	Hybrid	0.020	-0.017	0.197	0.474
	FHS	-0.497	-0.677	0.453	0.664
	XK_1	0.085	0.197	0.287	1.074
	XK_2	0.087	0.172	0.280	0.877
	CAViaR	0.067	0.045	0.180	0.261
	RiskM	-1.620	-1.811	1.161	0.976
500	Hybrid	0.014	0.027	0.080	0.071
	FHS	-0.366	-0.327	0.246	0.231
	XK_1	0.027	-0.021	0.100	0.214
	XK_2	0.030	0.000	0.099	0.214
	CAViaR	0.025	0.014	0.077	0.082
	RiskM	-1.590	-1.570	0.864	0.699
1000	Hybrid	0.001	-0.025	0.035	0.026
	FHS	-0.326	-0.340	0.207	0.162
	XK_1	0.008	-0.010	0.065	0.036
	XK_2	0.011	-0.003	0.071	0.038
	CAViaR	0.008	-0.008	0.038	0.028
	RiskM	-1.546	-1.538	0.825	0.786

Models 1 and 2. As shown in Table S.5, the comparative performance of the methods is quite different from that in Tables S.3 and S.4. The FHS method is worse than all the quantile regression based methods in terms of both the bias and MSE. The proposed method, together with the CAViaR method which has comparable performance, is superior to the other methods. This suggests that the proposed hybrid method indeed enjoys greater flexibility, and hence greater robustness, than the FHS method when the data have a more complex conditional quantile structure which cannot be captured by the global GARCH model.

2.2 Finite-sample comparison of efficiency with the unweighted estimator

To compare the efficiency of $\widehat{\theta}_{\tau n}$ and $\breve{\theta}_{\tau n}$ in finite samples, we generate the data from the GARCH$(1,1)$ model in (S.1) with standard normal or standardized Student's t_5-distributed innovations, using $(\alpha_0, \alpha_1, \beta_1) = (0.4, 0.2, 0.2)$ and $(\alpha_0, \alpha_1, \beta_1) = (0.4, 0.2, 0.6)$. The sample size is $n = 2000$, and two quantile levels, $\tau = 0.05$ and 0.1, are considered. Figure S.1 provides the box plots for the two estimators based on 1000 replications. It shows that the interquartile range of the weighted estimator $\widehat{\theta}_{\tau n}$ is smaller than that of the unweighted counterpart $\breve{\theta}_{\tau n}$ under all settings; the latter also suffers from more severe outliers. The efficiency gains from the weights seem larger for the Student's t_5 cases. Moreover, for the unweighted estimator $\breve{\theta}_{\tau n}$, the sample median slightly deviates from the true value $\theta_{\tau 0}$ especially when the innovations are Student's t_5-distributed. The results suggest that the weighted estimator is more efficient in finite samples.

2.3 Performance of the mixed bootstrapping procedure for more choices of random weights

In the last two experiments in Section 5, due to the limit of space, we only reported the results for standard exponential random weights. In this subsection, we provide the corresponding results for other choices of random weights.

In the second experiment in Section 5, we considered the residual QACF $r_{k,\tau}$ and the bootstrapping approximation of its asymptotic distribution. The data were generated from the GARCH$(1,1)$ model in (S.1) with $(\alpha_0, \alpha_1, \beta_1) = (0.1, 0.15, 0.8)$ and $\{\eta_t\}$ following the standard normal or standardized Student's t_5 distributions with unit variance.

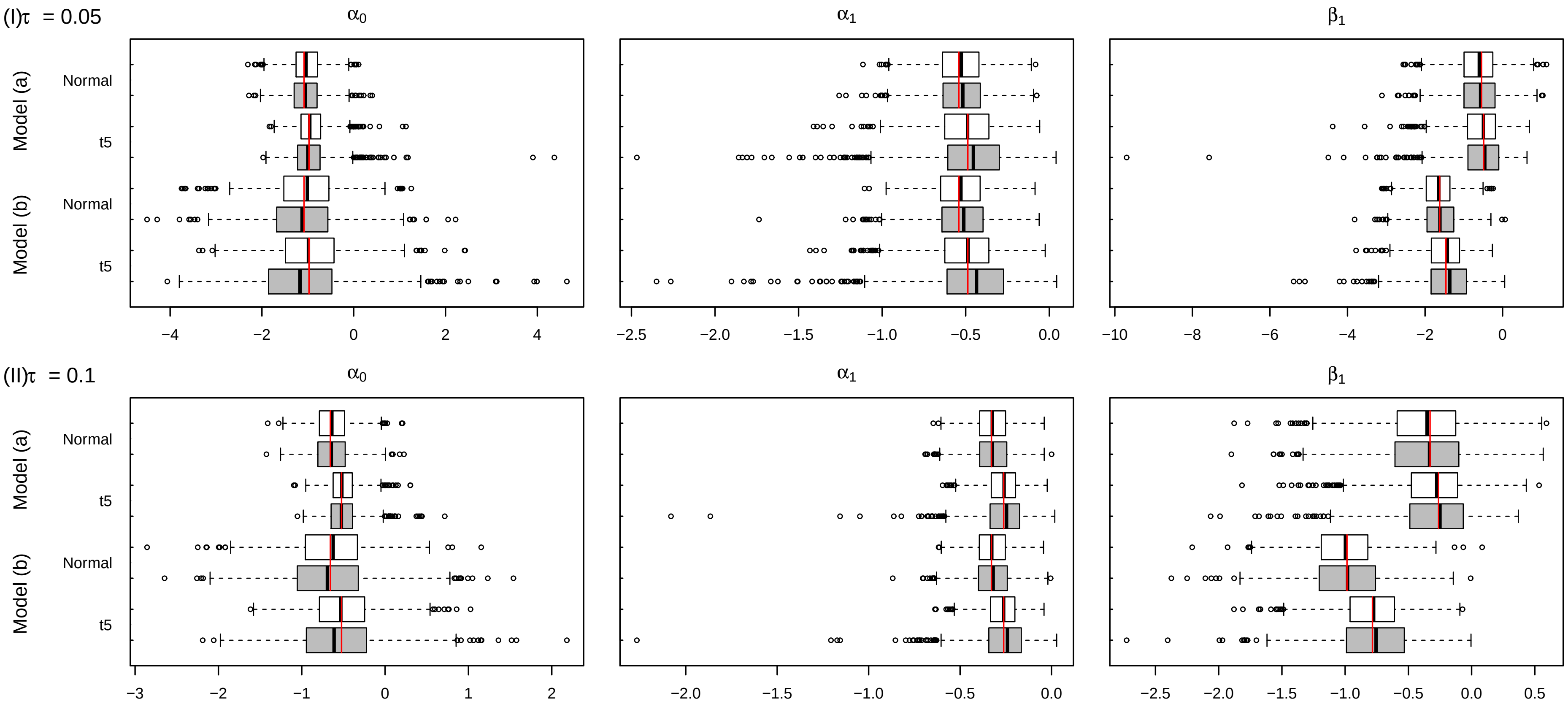

Figure S.1: Box plots for the weighted estimator $\widehat{\theta}_{\tau n}$ (white boxes) and the unweighted estimator $\breve{\theta}_{\tau n}$ (grey boxes), at $\tau = 0.05$ or 0.1, for two models with normal or Student's t_5-distributed innovations. Model (a): $(\alpha_0, \alpha_1, \beta_1) = (0.4, 0.2, 0.2)$; Model (b): $(\alpha_0, \alpha_1, \beta_1) = (0.4, 0.2, 0.6)$. The thick black line in the center of the box indicates the sample median, and the thin red line indicates the value of the corresponding element of the true parameter vector $\theta_{\tau 0}$. The notations α_0, α_1 and β_1 represent the corresponding elements of $\widehat{\theta}_{\tau n}$ and $\breve{\theta}_{\tau n}$.

Table S.6: Bias (×100), ESD (×100) and ASD (×100) for the residual QACF $r_{k,\tau}$ at $\tau = 0.05$ or 0.1 and $k = 2, 4$ or 6, for normal or Student's t_5-distributed innovations, where ASD_i corresponds to random weight W_i for $i = 1, \ldots, 4$.

		Normal distribution						Student's t_5 distribution					
n	k	Bias	ESD	ASD_1	ASD_2	ASD_3	ASD_4	Bias	ESD	ASD_1	ASD_2	ASD_3	ASD_4
							$\tau = 0.05$						
500	2	1.27	4.88	6.72	6.33	6.50	6.52	0.78	4.36	5.91	5.71	5.75	5.82
	4	0.90	4.88	6.83	6.44	6.62	6.63	0.69	4.67	5.94	5.73	5.79	5.84
	6	1.04	4.91	6.81	6.46	6.62	6.63	0.37	4.75	6.03	5.83	5.87	5.93
1000	2	0.48	3.24	4.05	3.90	3.97	3.97	0.30	3.13	3.57	3.53	3.53	3.55
	4	0.50	3.34	4.09	3.94	4.01	4.02	0.35	3.13	3.54	3.48	3.49	3.51
	6	0.43	3.29	4.13	3.99	4.06	4.06	0.18	3.35	3.66	3.61	3.62	3.64
2000	2	0.29	2.23	2.59	2.53	2.56	2.56	0.28	2.15	2.30	2.29	2.29	2.30
	4	0.15	2.26	2.62	2.56	2.59	2.59	0.10	2.26	2.31	2.30	2.30	2.30
	6	0.16	2.25	2.63	2.57	2.60	2.60	0.15	2.20	2.32	2.31	2.31	2.31
							$\tau = 0.1$						
500	2	0.67	4.35	5.34	5.22	5.27	5.28	0.69	4.32	4.82	4.87	4.83	4.85
	4	0.47	4.59	5.43	5.31	5.36	5.38	0.42	4.31	4.84	4.87	4.85	4.86
	6	0.61	4.64	5.44	5.33	5.37	5.39	0.08	4.52	4.90	4.93	4.90	4.92
1000	2	0.36	3.13	3.44	3.41	3.42	3.43	0.25	3.14	3.26	3.28	3.26	3.28
	4	0.15	3.19	3.51	3.47	3.49	3.49	0.30	3.01	3.17	3.19	3.18	3.18
	6	0.30	3.16	3.54	3.51	3.52	3.53	-0.01	3.20	3.29	3.30	3.29	3.30
2000	2	0.20	2.23	2.33	2.32	2.32	2.33	0.09	2.21	2.23	2.24	2.23	2.24
	4	0.02	2.14	2.36	2.35	2.36	2.36	0.10	2.19	2.21	2.21	2.20	2.21
	6	0.14	2.19	2.38	2.37	2.38	2.37	0.04	2.18	2.23	2.23	2.23	2.23

Table S.7: Rejection rate (%) of the test statistic $Q(K)$ for $K = 6$ at the 5% significance level, for normal or Student's t_5-distributed innovations and $d = 0$, 0.3 or 0.6, where Q_i denotes the test statistic based on random weight W_i for $i = 1, \ldots, 4$.

		Normal distribution				Student's t_5 distribution			
n	d	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
					$\tau = 0.05$				
500	0.0	2.8	2.0	2.1	2.0	1.9	1.8	2.1	1.4
	0.3	4.8	3.3	4.2	3.6	3.8	2.7	3.3	2.9
	0.6	7.4	6.9	6.7	6.8	7.8	7.2	8.4	7.3
1000	0.0	3.3	2.9	3.4	3.3	3.0	3.2	3.4	3.0
	0.3	7.2	6.5	6.6	6.5	10.6	9.8	10.7	9.9
	0.6	21.6	21.4	21.8	21.4	29.4	29.6	29.5	29.5
2000	0.0	4.5	4.4	4.4	4.4	5.3	4.7	5.2	4.8
	0.3	16.1	15.7	15.4	15.9	27.9	26.5	27.4	26.9
	0.6	55.2	54.8	54.7	55.0	69.8	70.6	69.8	70.6
					$\tau = 0.1$				
500	0.0	3.4	2.9	3.1	3.2	3.4	3.7	3.7	3.2
	0.3	6.9	6.0	6.8	6.4	6.5	5.9	6.2	5.9
	0.6	27.0	25.1	26.7	25.8	21.0	20.1	20.1	20.1
1000	0.0	4.0	3.7	3.9	4.4	4.3	4.2	4.3	3.9
	0.3	15.7	15.8	15.5	15.4	16.3	15.7	16.5	16.0
	0.6	60.9	61.3	61.0	60.7	46.8	47.2	46.6	46.9
2000	0.0	4.9	4.4	4.5	4.5	4.3	4.2	4.2	4.2
	0.3	36.5	35.5	36.0	36.3	34.3	34.2	33.8	33.4
	0.6	92.5	93.3	92.9	92.8	83.2	83.1	82.9	83.1

Three sample sizes, $n = 500$, 1000 and 2000, and two quantile levels, $\tau = 0.05$ and 0.1, were considered, with 1000 replications for each sample size. Table S.6 reports the results for four distributions for the random weights $\{\omega_t\}$: the standard exponential distribution (W_1); the Rademacher distribution (W_2), which takes the values 0 or 2, each with probability 0.5 (Li et al., 2014); Mammen's two-point distribution (W_3), which takes the value $(-\sqrt{5}+3)/2$ with probability $(\sqrt{5}+1)/2\sqrt{5}$ and the value $(\sqrt{5}+3)/2$ with probability $1-(\sqrt{5}+1)/2\sqrt{5}$ (Mammen, 1993); and a mixture of the standard exponential distribution and the Rademacher distribution (W_4) with mixing probability 0.5. We can observe that the choice of random weights has little influence on the bootstrap approximations, and our other findings are similar to those in Section 5.

In the third experiment in Section 5, we evaluated the empirical size and power of the proposed test statistic $Q(K)$. The data generating process was

$$x_t = \sqrt{h_t}\eta_t, \quad h_t = 0.4 + 0.2x_{t-1}^2 + dx_{t-4}^2 + 0.2h_{t-1},$$

with $\{\eta_t\}$ following the same distributions as in the previous experiment, and the departure $d = 0$, 0.3 or 0.6. We conducted the estimation assuming a GARCH$(1,1)$ model; thus, $d = 0$ corresponds to the size of the test, and $d \neq 0$ corresponds to the power. Table S.7 reports the results for the four random weights distributions considered in the previous experiment. Again, the performance is insensitive to the choice of random weights, and our other findings are also similar to those reported in Section 5.

3 Additional results for the empirical analysis

3.1 Case-by-case comparison with the FHS method

To further illustrate the superiority of the proposed method over the FHS method in the empirical analysis, we conduct a case-by-case comparison of the two methods based on the results reported in Table 4 of the paper.

As the CC and DQ tests are complementary to each other, we focus on the minimum of the two p-values to evaluate the backtesting performance. We categorize the 18 cases reported in Table 4 in the paper (i.e., six quantile levels for three stock market indexes) into three groups and employ the following criteria for each group:

- Category-0: If for both methods, the minimum p-value is less than 0.05 (i.e., at least one p-value is less than 0.05), then both methods are equally poor, and hence we do not compare them for this case.

- Category-1: If for both methods, the minimum p-value is larger than 0.2 (i.e., both p-values are larger than 0.2), then both methods are equally good in terms of backtesting, and the method with the smaller empirical coverage error is better.

- Category-2: For all other cases, the method with the larger minimum p-value is better; i.e., the backtesting result determines which method is better.

Table S.8 summarizes the empirical coverage error, minimum p-values, the winner and the corresponding category of each case, which shows that the proposed hybrid

Table S.8: Empirical coverage error (%), minimum p-value of the CC and DQ tests, the better method (H, Hybrid; F, FHS) and corresponding category for all the 18 cases considered in the paper.

	ECE		Minimum p-value		Better method
	Hybrid	FHS	Hybrid	FHS	(category)
			S&P 500		
L1.0	-0.02	0.04	0.000	0.082	F (2)
L2.5	-0.48	-0.36	0.001	0.005	-
L5.0	-0.90	-1.15	0.017	0.016	-
U5.0	0.54	0.84	0.245	0.244	H (1)
U2.5	0.30	0.42	0.356	0.222	H (1)
U1.0	0.08	0.33	0.275	0.342	H (1)
			Dow 30		
L1.0	-0.14	0.16	0.063	0.115	F (2)
L2.5	-0.54	-0.24	0.000	0.000	-
L5.0	-0.72	-0.78	0.000	0.027	-
U5.0	0.84	1.21	0.273	0.064	H (2)
U2.5	0.11	0.24	0.568	0.806	H (1)
U1.0	-0.28	0.39	0.418	0.221	H (1)
			HSI		
L1.0	0.11	-0.02	0.393	0.425	F (1)
L2.5	-0.04	0.14	0.362	0.290	H (1)
L5.0	-0.69	-0.69	0.421	0.159	H (2)
U5.0	0.14	-0.23	0.766	0.635	H (1)
U2.5	0.04	0.35	0.477	0.631	H (1)
U1.0	-0.17	0.26	0.048	0.492	F (2)

method is better for 10 cases, while the FHS method is better for 4 cases. Therefore, we may conclude that the proposed method achieves better performance overall.

3.2 Performance of the proposed method after quantile rearrangements

We have also re-calculated both the backtesting and empirical coverage results for the proposed hybrid method after conducting the quantile rearrangement in Chernozhukov et al. (2010). We find that all empirical coverage errors do not change at all from the results in Table 4 in the paper, after the percentage points are rounded down to two decimal places. Moreover, the p-values of the CC and DQ tests change very little; see Table S.9 for the p-values of the tests for the proposed method before and after the quantile rearrangement for all the 18 cases considered in the paper.

Table S.9: p-values of the CC and DQ tests for the proposed method before and after the quantile rearrangement for all the 18 cases considered in the paper.

		Before rearrangement			After rearrangement		
τ		S&P 500	Dow 30	HSI	S&P 500	Dow 30	HSI
L1.0	CC	0.031	0.228	0.746	0.031	0.228	0.746
	DQ	0.000	0.063	0.393	0.000	0.064	0.397
L2.5	CC	0.007	0.028	0.362	0.007	0.028	0.362
	DQ	0.001	0.000	0.571	0.001	0.000	0.570
L5.0	CC	0.017	0.212	0.421	0.017	0.212	0.421
	DQ	0.017	0.000	0.556	0.017	0.000	0.558
U5.0	CC	0.544	0.273	0.866	0.544	0.273	0.866
	DQ	0.245	0.281	0.766	0.245	0.282	0.763
U2.5	CC	0.356	0.568	0.995	0.356	0.568	0.995
	DQ	0.585	0.656	0.477	0.585	0.655	0.471
U1.0	CC	0.275	0.418	0.640	0.275	0.418	0.640
	DQ	0.384	0.905	0.048	0.384	0.906	0.054

4 Technical details

This section gives detailed proofs of Lemma A.1, Theorems 1–4, Corollaries 1–3 and Equation (2.6). A preliminary Lemma S.1 is also included, which is used to prove the asymptotic negligibility of the effect of the initial values $\{x_0^2, \dots, x_{1-q}^2, h_0, \dots, h_{1-p}\}$.

Throughout the proofs, C is a generic positive constant which may take different values at its different occurrences, and $C(M)$ is such a constant whose value depends on M. We denote by $\|\cdot\|$ the norm of a matrix or column vector, defined as $\|A\| = \sqrt{\text{tr}(AA')} = \sqrt{\sum_{i,j} |a_{ij}|^2}$. In addition, let $z_t(\theta) = (1, x_{t-1}^2, \dots, x_{t-q}^2, h_{t-1}(\theta), \dots, h_{t-p}(\theta))'$, $\widetilde{z}_t(\theta) = (1, x_{t-1}^2, \dots, x_{t-q}^2, \widetilde{h}_{t-1}(\theta), \dots, \widetilde{h}_{t-p}(\theta))'$, and, for simplicity, write $z_t = z_t(\theta_0)$, $\breve{z}_t = \widetilde{z}_t(\theta_0)$, and $\widetilde{z}_t = \widetilde{z}_t(\widetilde{\theta}_n)$, where $\widetilde{\theta}_n$ is the Gaussian QMLE of model (1.1). In the proofs of Theorems 2 and 4, the notations E^*, $O_p^*(1)$ and $o_p^*(1)$ correspond to the bootstrap probability space.

Lemma S.1. *Under Assumption 1,*

$$\sup_{\theta\in\Theta}|\widetilde{h}_t(\theta)-h_t(\theta)|\leqslant C\rho^t\zeta \quad \textit{and} \quad \sup_{\theta\in\Theta}\left\|\frac{\partial\widetilde{h}_t(\theta)}{\partial\theta}-\frac{\partial h_t(\theta)}{\partial\theta}\right\|\leqslant C\rho^t\zeta,$$

where $C>0$ and $0<\rho<1$ are constants, and ζ is a random variable independent of t with $E|\zeta|^{\delta_0}<\infty$ for some $\delta_0>0$.

Proof of Lemma S.1. The lemma can be proved by a method similar to that for Equations (6) and (7) in the proof of Theorem 1 in Zheng et al. (2016). □

Proof of Lemma A.1. We first prove (i). For any $\theta=(\alpha_0,\alpha_1,\ldots,\alpha_q,\beta_1,\ldots,\beta_p)'\in\Theta$ and $\gamma>1$, define

$$U(\gamma,\theta)=\{\theta^*=(\alpha_0^*,\alpha_1^*,\ldots,\alpha_q^*,\beta_1^*,\ldots,\beta_p^*)'\in\Theta:\max_{1\leqslant j\leqslant p}\frac{\beta_j^*}{\beta_j}\leqslant\gamma\}.$$

Claim (i) follows from a more general result: for any $\kappa>0$, there is $\gamma>1$ such that

$$E\left[\sup_{\theta\in\Theta}\sup_{\theta^*\in U(\gamma,\theta)}\frac{h_t(\theta^*)}{h_t(\theta)}\right]^\kappa<\infty. \tag{S.2}$$

Notice that for any θ, the set $U(\gamma,\theta)$ only imposes an upper bound on the β_j^*'s, while the condition $\|\theta_1-\theta_2\|\leqslant c$ restricts the distance between θ_1 and θ_2.

We shall prove (S.2). Note that the functions $h_t(\theta)$, as defined recursively in (2.2), can be written in the form of

$$h_t(\theta)=c_0(\theta)+\sum_{j=1}^{\infty}c_j(\theta)x_{t-j}^2,$$

and the series converges with probability one for all $\theta\in\Theta$; see, e.g., Berkes et al. (2003). Moreover, $c_0(\theta)=\alpha_0/(1-\beta_1-\cdots-\beta_p)\geqslant C_1=\underline{w}/(1-p\underline{w})>0$ for all $\theta\in\Theta$, and from Lemma 3.1 in Berkes et al. (2003), it holds that

$$\sup_{\theta\in\Theta}c_j(\theta)\leqslant C_2\rho_1^j, \quad j\geqslant 0, \tag{S.3}$$

where $\rho_1=\rho_0^{1/p}\in(0,1)$, and

$$\sup_{\theta\in\Theta}\sup_{\theta^*\in U(\gamma,\theta)}\frac{c_j(\theta^*)}{c_j(\theta)}\leqslant C_3\gamma^j, \quad j\geqslant 0, \tag{S.4}$$

for some constants $C_2,C_3>0$. Using (S.4), we have

$$\sup_{\theta\in\Theta}\sup_{\theta^*\in U(\gamma,\theta)}\frac{h_t(\theta^*)}{h_t(\theta)}\leqslant\frac{C_2}{C_1}+C_3\sup_{\theta\in\Theta}\frac{\sum_{j=1}^{\infty}\gamma^jc_j(\theta)x_{t-j}^2}{C_1+\sum_{j=1}^{\infty}c_j(\theta)x_{t-j}^2},$$

and then it suffices to show that for any $\kappa \geqslant 1$,

$$\left\| \sup_{\theta \in \Theta} \frac{\sum_{j=1}^{\infty} \gamma^j c_j(\theta) x_{t-j}^2}{C_1 + \sum_{j=1}^{\infty} c_j(\theta) x_{t-j}^2} \right\|_\kappa < \infty,$$

where $\|\cdot\|_\kappa$ denotes the L_κ norm, i.e., $\|X\|_\kappa = (E|X|^\kappa)^{1/\kappa}$. Note that there is $\delta_0 > 0$ such that $E|x_0^2|^{\delta_0} < \infty$. Thus, for any $\kappa \geqslant 1$ and $\delta_1 \in (1 - \delta_0/\kappa, 1)$, by (S.3) and the Minkowski inequality, we have

$$\begin{aligned}
\left\| \sup_{\theta \in \Theta} \frac{\sum_{j=1}^{\infty} \gamma^j c_j(\theta) x_{t-j}^2}{C_1 + \sum_{j=1}^{\infty} c_j(\theta) x_{t-j}^2} \right\|_\kappa &\leqslant \left\| \sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \frac{\gamma^j c_j(\theta) x_{t-j}^2}{C_1^{1-\delta_1} [c_j(\theta) x_{t-j}^2]^{\delta_1}} \right\|_\kappa \\
&\leqslant C_1^{-(1-\delta_1)} \left\| \sum_{j=1}^{\infty} \gamma^j (C_2 \rho_1^j x_{t-j}^2)^{1-\delta_1} \right\|_\kappa \\
&\leqslant C \sum_{j=1}^{\infty} (\gamma \rho_1^{1-\delta_1})^j \left[E|x_0^2|^{(1-\delta_1)\kappa} \right]^{1/\kappa} < \infty,
\end{aligned}$$

if γ is close enough to 1. Therefore, (S.2) holds, and so does (i).

From the proof of Theorem 2.2 in Francq and Zakoian (2004), under Assumption 1, for any $\kappa > 0$,

$$E \sup_{\theta \in \Theta} \left\| \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \right\|^\kappa < \infty, \quad E \sup_{\theta \in \Theta} \left\| \frac{1}{h_t(\theta)} \frac{\partial^2 h_t(\theta)}{\partial \theta \partial \theta'} \right\|^\kappa < \infty \quad \text{and}$$

$$E \sup_{\theta \in \Theta} \left| \frac{1}{h_t(\theta)} \frac{\partial^3 h_t(\theta)}{\partial \theta_i \partial \theta_k \partial \theta_\ell} \right|^\kappa < \infty,$$

where $1 \leqslant i, k, \ell \leqslant p+q+1$; see also Lemma 3.6 in Berkes and Horváth (2004). Combining these with (i), we immediately obtain (ii)-(iv). □

Proof of Theorem 1. Let $L_n(\theta) = \sum_{t=1}^n \tilde{h}_t^{-1} \rho_\tau(y_t - \theta' \tilde{z}_t)$ and $\breve{L}_n(\theta) = \sum_{t=1}^n \tilde{h}_t^{-1} \rho_\tau(y_t - \theta' \breve{z}_t)$. Notice that for $x \neq 0$,

$$\rho_\tau(x - y) - \rho_\tau(x) = -y\psi_\tau(x) + \int_0^y [I(x \leqslant s) - I(x \leqslant 0)] ds, \tag{S.5}$$

where $\psi_\tau(x) = \tau - I(x < 0)$; see Knight (1998). Then, for any fixed $u \in \mathbb{R}^{p+q+1}$,

$$L_n(\theta_{\tau 0} + n^{-1/2} u) - \breve{L}_n(\theta_{\tau 0}) = -L_{1n}(u) + L_{2n}(u), \tag{S.6}$$

where

$$\begin{aligned}
L_{1n}(u) &= \sum_{t=1}^n \psi_\tau(\breve{e}_{t,\tau}) \tilde{h}_t^{-1} \left[(\theta_{\tau 0} + n^{-1/2} u)' \tilde{z}_t - \theta_{\tau 0}' \breve{z}_t \right], \\
L_{2n}(u) &= \sum_{t=1}^n \tilde{h}_t^{-1} \int_0^{(\theta_{\tau 0} + n^{-1/2} u)' \tilde{z}_t - \theta_{\tau 0}' \breve{z}_t} \left[I(\breve{e}_{t,\tau} \leqslant s) - I(\breve{e}_{t,\tau} \leqslant 0) \right] ds,
\end{aligned}$$

and $\breve{e}_{t,\tau} = y_t - \theta_{\tau 0}'\breve{z}_t$. Let $u^{(j)}$ be the $(j+q+1)$-th element of u, and denote $\beta_{\tau 0}^{(j)} = b_\tau \beta_{0j}$, for $j = 1, \ldots, p$. It can be verified that

$$(\theta_{\tau 0} + n^{-1/2}u)'\widetilde{z}_t - \theta_{\tau 0}'\breve{z}_t = \xi_{1nt}(\widetilde{\theta}_n) + \xi_{2nt}(\widetilde{\theta}_n) + \xi_{3nt}(\widetilde{\theta}_n), \tag{S.7}$$

where

$$\begin{aligned}
\xi_{1nt}(\theta) &= n^{-1/2}u'z_t + \sum_{j=1}^p \beta_{\tau 0}^{(j)} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta'}(\theta - \theta_0), \\
\xi_{2nt}(\theta) &= \frac{1}{\sqrt{n}} \sum_{j=1}^p u^{(j)}[h_{t-j}(\theta) - h_{t-j}] + \sum_{j=1}^p \beta_{\tau 0}^{(j)} \left[h_{t-j}(\theta) - h_{t-j} - \frac{\partial h_{t-j}(\theta_0)}{\partial \theta'}(\theta - \theta_0) \right], \\
\xi_{3nt}(\theta) &= \frac{1}{\sqrt{n}} \sum_{j=1}^p u^{(j)}[\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)] \\
&\quad + \sum_{j=1}^p \beta_{\tau 0}^{(j)} \left\{ [\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)] - [\widetilde{h}_{t-j}(\theta_0) - h_{t-j}] \right\}.
\end{aligned}$$

For any $M > 0$, denote $\Theta_n = \Theta_n(M) = \{\theta \in \Theta : \|\theta - \theta_0\| \leqslant n^{-1/2}M\}$. Using the Taylor expansion, it holds that

$$\sup_{\theta \in \Theta_n} |\xi_{2nt}(\theta)| \leqslant \frac{M}{n} \sum_{j=1}^p |u^{(j)}| \sup_{\theta \in \Theta_n} \left\| \frac{\partial h_{t-j}(\theta)}{\partial \theta} \right\| + \frac{M^2}{2n} \sum_{j=1}^p |\beta_{\tau 0}^{(j)}| \sup_{\theta \in \Theta_n} \left\| \frac{\partial^2 h_{t-j}(\theta)}{\partial \theta \partial \theta'} \right\|, \tag{S.8}$$

and by Lemma S.1,

$$\begin{aligned}
&\sup_{\theta \in \Theta_n} |\xi_{3nt}(\theta)| \\
&\quad \leqslant \frac{1}{\sqrt{n}} \sum_{j=1}^p \left[|u^{(j)}| \sup_{\theta \in \Theta} |\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)| + M|\beta_{\tau 0}^{(j)}| \sup_{\theta \in \Theta} \left\| \frac{\partial \widetilde{h}_{t-j}(\theta)}{\partial \theta} - \frac{\partial h_{t-j}(\theta)}{\partial \theta} \right\| \right] \\
&\quad \leqslant n^{-1/2}C(M)\rho^t \zeta.
\end{aligned} \tag{S.9}$$

Moreover,

$$\breve{e}_{t,\tau} = (\varepsilon_t - b_\tau)h_t + a_t, \quad \text{where} \quad a_t = \sum_{j=1}^p \beta_{\tau 0}^{(j)} [h_{t-j} - \widetilde{h}_{t-j}(\theta_0)] \in \mathcal{F}_0. \tag{S.10}$$

We first consider $L_{1n}(u)$, which can be decomposed into four parts,

$$L_{1n}(u) = \sum_{t=1}^n A_{1nt}(\widetilde{\theta}_n) + \sum_{t=1}^n A_{2nt}(\widetilde{\theta}_n) + \sum_{t=1}^n A_{3nt}(\widetilde{\theta}_n) + \sum_{t=1}^n A_{4nt}(\widetilde{\theta}_n), \tag{S.11}$$

where

$$A_{1nt}(\theta) = \psi_\tau(\breve{e}_{t,\tau})\widetilde{h}_t^{-1}(\theta)\xi_{3nt}(\theta) + \psi_\tau(\breve{e}_{t,\tau})[\widetilde{h}_t^{-1}(\theta) - h_t^{-1}(\theta)][\xi_{1nt}(\theta) + \xi_{2nt}(\theta)],$$

$$A_{2nt}(\theta) = [\psi_\tau(\breve{e}_{t,\tau}) - \psi_\tau(\varepsilon_t - b_\tau)]h_t^{-1}(\theta)[\xi_{1nt}(\theta) + \xi_{2nt}(\theta)],$$

$$A_{3nt}(\theta) = \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}(\theta)\xi_{2nt}(\theta) \quad \text{and} \quad A_{4nt}(\theta) = \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}(\theta)\xi_{1nt}(\theta).$$

Note that $\inf_{\theta\in\Theta} h_t(\theta) \geqslant \underline{w}$ and $\inf_{\theta\in\Theta} \widetilde{h}_t(\theta) \geqslant \underline{w}$. By Lemma S.1, (S.8) and (S.9), we can show that

$$\begin{aligned}\sup_{\theta\in\Theta_n}\left|\sum_{t=1}^n A_{1nt}(\theta)\right| &\leqslant \frac{1}{\underline{w}}\sum_{t=1}^n \sup_{\theta\in\Theta_n}|\xi_{3nt}(\theta)| + \frac{C\zeta}{\underline{w}^2}\sum_{t=1}^n \rho^t \sup_{\theta\in\Theta_n}(|\xi_{1nt}(\theta)| + |\xi_{2nt}(\theta)|) \\ &= o_p(1), \end{aligned} \tag{S.12}$$

which, together with the fact that $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, implies that

$$\sum_{t=1}^n A_{1nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.13}$$

Note that by Lemma S.1 and Assumption 2, we have

$$|F(b_\tau) - F(b_\tau - h_t^{-1}a_t)| \leqslant \sup_{x\in\mathbb{R}} f(x)\sum_{j=1}^p \frac{|\beta_{\tau 0}^{(j)}|}{\underline{w}}|h_{t-j}(\theta_0) - \widetilde{h}_{t-j}(\theta_0)| \leqslant C\rho^t\zeta.$$

It then follows from (S.10) that

$$\begin{aligned} E|\psi_\tau(\breve{e}_{t,\tau}) - \psi_\tau(\varepsilon_t - b_\tau)| &= E|F(b_\tau) - F(b_\tau - h_t^{-1}a_t)| \\ &= E[|F(b_\tau) - F(b_\tau - h_t^{-1}a_t)|I(C\rho^t\zeta \leqslant \rho^{t/2})] \\ &\quad + E[|F(b_\tau) - F(b_\tau - h_t^{-1}a_t)|I(C\rho^t\zeta > \rho^{t/2})] \\ &\leqslant \rho^{t/2} + \Pr(C\rho^t\zeta > \rho^{t/2}) \leqslant \rho^{t/2} + C\rho^{\delta_0 t/2}, \end{aligned} \tag{S.14}$$

where we used the Markov inequality and the fact that $E|\zeta|^{\delta_0} < \infty$. Moreover,

$$\|h_t^{-1}z_t\| \leqslant \frac{\sqrt{p+q+1}}{\underline{w}}, \tag{S.15}$$

$$\sup_{\theta_1,\theta_2\in\Theta_n}\left|\frac{\xi_{1nt}(\theta_2)}{h_t(\theta_1)}\right| \leqslant \frac{|h_t^{-1}u'z_t|}{\sqrt{n}}\sup_{\theta\in\Theta_n}\frac{h_t}{h_t(\theta)} + \frac{M}{\underline{w}\sqrt{n}}\sum_{j=1}^p|\beta_{\tau 0}^{(j)}|\sup_{\theta\in\Theta_n}\left\|\frac{1}{h_{t-j}(\theta)}\frac{\partial h_{t-j}(\theta_0)}{\partial\theta}\right\|, \tag{S.16}$$

and by the Taylor expansion,

$$\begin{aligned}\sup_{\theta_1,\theta_2\in\Theta_n}\left|\frac{\xi_{2nt}(\theta_2)}{h_t(\theta_1)}\right| &\leqslant \frac{M}{\underline{w}n}\sum_{j=1}^p|u^{(j)}|\sup_{\theta_1,\theta_2\in\Theta_n}\left\|\frac{1}{h_{t-j}(\theta_1)}\frac{\partial h_{t-j}(\theta_2)}{\partial\theta}\right\| \\ &\quad + \frac{M^2}{2\underline{w}n}\sum_{j=1}^p|\beta_{\tau 0}^{(j)}|\sup_{\theta_1,\theta_2\in\Theta_n}\left\|\frac{1}{h_{t-j}(\theta_1)}\frac{\partial^2 h_{t-j}(\theta_2)}{\partial\theta\partial\theta'}\right\|. \end{aligned} \tag{S.17}$$

As a result, by the Hölder inequality, Lemma A.1 and (S.14)-(S.17), we have

$$\begin{aligned} E\sup_{\theta\in\Theta_n}\left|\sum_{t=1}^n A_{2nt}(\theta)\right| &\leqslant \sum_{t=1}^n [E|\psi_\tau(\breve{e}_{t,\tau}) - \psi_\tau(\varepsilon_t - b_\tau)|]^{1/2}\left[E\sup_{\theta\in\Theta_n}\left(\frac{|\xi_{1nt}(\theta)| + |\xi_{2nt}(\theta)|}{h_t(\theta)}\right)^2\right]^{1/2} \\ &= o(1), \end{aligned}$$

which, together with the fact that $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, implies that

$$\sum_{t=1}^{n} A_{2nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.18}$$

Applying the Taylor expansion to $h_t^{-1}(\theta)$ and $\xi_{2nt}(\theta)$ respectively, we have

$$h_t^{-1}(\theta)\xi_{2nt}(\theta) = \xi_{4nt}(\theta) + \xi_{5nt}(\theta), \tag{S.19}$$

where

$$\begin{aligned}
\xi_{4nt}(\theta) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{p} \frac{u^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta'}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \sum_{j=1}^{p} \frac{\beta_{\tau 0}^{(j)}}{h_t} \frac{\partial^2 h_{t-j}(\theta_0)}{\partial \theta \partial \theta'}(\theta - \theta_0), \\
\xi_{5nt}(\theta) &= -\frac{\xi_{2nt}(\theta)}{h_t^2(\theta_1^*)} \frac{\partial h_t(\theta_1^*)}{\partial \theta'}(\theta - \theta_0) + \frac{(\theta - \theta_0)'}{2\sqrt{n}} \sum_{j=1}^{p} \frac{u^{(j)}}{h_t} \frac{\partial^2 h_{t-j}(\theta_2^*)}{\partial \theta \partial \theta'}(\theta - \theta_0) \\
&\quad + \frac{1}{6} \sum_{j=1}^{p} \sum_{i,k,\ell=1}^{p+q+1} \frac{\beta_{\tau 0}^{(j)}}{h_t} \frac{\partial^3 h_{t-j}(\theta_2^*)}{\partial \theta_i \partial \theta_k \partial \theta_\ell}(\theta_i - \theta_{0i})(\theta_k - \theta_{0k})(\theta_\ell - \theta_{0\ell}),
\end{aligned}$$

with θ_1^* and θ_2^* both between θ and θ_0. Then, it follows from Lemma A.1, the ergodic theorem and $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$ that

$$\sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)\xi_{4nt}(\widetilde{\theta}_n) = o_p(1) \tag{S.20}$$

and

$$E \sup_{\theta \in \Theta_n} \left| \sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)\xi_{5nt}(\theta) \right| \leqslant \sum_{t=1}^{n} E \sup_{\theta \in \Theta_n} |\xi_{5nt}(\theta)| = O(n^{-1/2}), \tag{S.21}$$

which implies

$$\sum_{t=1}^{n} A_{3nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.22}$$

By a method similar to that for $\sum_{t=1}^{n} A_{3nt}(\widetilde{\theta}_n)$, we can show that

$$\sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)[h_t^{-1}(\widetilde{\theta}_n) - h_t^{-1}]\xi_{1nt}(\widetilde{\theta}_n) = o_p(1),$$

which implies

$$\sum_{t=1}^{n} A_{4nt}(\widetilde{\theta}_n) = \sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}\xi_{1nt}(\widetilde{\theta}_n) + o_p(1) = u'T_{1n} + T_{2n} + o_p(1), \tag{S.23}$$

where

$$T_{1n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau) \frac{z_t}{h_t} \text{ and } T_{2n} = \sqrt{n}(\widetilde{\theta}_n - \theta_0)' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(\varepsilon_t - b_\tau) \sum_{j=1}^{p} \frac{\beta_{\tau 0}^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta}.$$

Combining (S.11), (S.13), (S.18), (S.22), and (S.23), we have

$$L_{1n}(u) = u'T_{1n} + T_{2n} + o_p(1). \tag{S.24}$$

Next we consider $L_{2n}(u)$. For simplicity, denote $I_t^*(s) = I(\breve{e}_{t,\tau} \leqslant s) - I(\breve{e}_{t,\tau} \leqslant 0)$. From (S.7), we have the decomposition

$$L_{2n}(u) = \sum_{t=1}^{n} B_{1nt}(\widetilde{\theta}_n) + \sum_{t=1}^{n} B_{2nt}(\widetilde{\theta}_n) + \sum_{t=1}^{n} B_{3nt}(\widetilde{\theta}_n) + \sum_{t=1}^{n} B_{4nt}(\widetilde{\theta}_n), \tag{S.25}$$

where

$$\begin{aligned}
B_{1nt}(\theta) &= \widetilde{h}_t^{-1}(\theta) \int_{\xi_{1nt}(\theta)+\xi_{2nt}(\theta)}^{\xi_{1nt}(\theta)+\xi_{2nt}(\theta)+\xi_{3nt}(\theta)} I_t^*(s)ds + [\widetilde{h}_t^{-1}(\theta) - h_t^{-1}(\theta)] \int_0^{\xi_{1nt}(\theta)+\xi_{2nt}(\theta)} I_t^*(s)ds, \\
B_{2nt}(\theta) &= h_t^{-1}(\theta) \int_{\xi_{1nt}(\theta)}^{\xi_{1nt}(\theta)+\xi_{2nt}(\theta)} I_t^*(s)ds, \\
B_{3nt}(\theta) &= [h_t^{-1}(\theta) - h_t^{-1}] \int_0^{\xi_{1nt}(\theta)} I_t^*(s)ds, \quad \text{and} \quad B_{4nt}(\theta) = h_t^{-1} \int_0^{\xi_{1nt}(\theta)} I_t^*(s)ds.
\end{aligned}$$

By a method similar to that for (S.13), we can show that

$$\begin{aligned}
\sup_{\theta\in\Theta_n} \left| \sum_{t=1}^{n} B_{1nt}(\theta) \right| &\leqslant \sum_{t=1}^{n} \sup_{\theta\in\Theta_n} \left[\frac{|\xi_{3nt}(\theta)|}{\widetilde{h}_t(\theta)} + \left| \frac{1}{\widetilde{h}_t(\theta)} - \frac{1}{h_t(\theta)} \right| (|\xi_{1nt}(\theta)| + |\xi_{2nt}(\theta)|) \right] \\
&= o_p(1),
\end{aligned} \tag{S.26}$$

which, together with the fact that $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, implies

$$\sum_{t=1}^{n} B_{1nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.27}$$

From (S.10), (S.16), (S.17), Assumption 2 and the Hölder inequality, we have

$$\begin{aligned}
&E \sup_{\theta\in\Theta_n} \left| \sum_{t=1}^{n} B_{2nt}(\theta) \right| \\
&\quad \leqslant E \sum_{t=1}^{n} \sup_{\theta\in\Theta_n} |h_t^{-1}(\theta)\xi_{2nt}(\theta)| I \left(|\breve{e}_{t,\tau}| \leqslant \sup_{\theta\in\Theta_n} (|\xi_{1nt}(\theta)| + |\xi_{2nt}(\theta)|) \right) \\
&\quad \leqslant \sqrt{2 \sup_{x\in\mathbb{R}} f(x)} \sum_{t=1}^{n} \left[E \sup_{\theta\in\Theta_n} \left| \frac{\xi_{2nt}(\theta)}{h_t(\theta)} \right|^2 \right]^{1/2} \left[E \sup_{\theta\in\Theta_n} \frac{(|\xi_{1nt}(\theta)| + |\xi_{2nt}(\theta)|)}{h_t} \right]^{1/2} \\
&\quad = o(1),
\end{aligned}$$

which, combined with the fact that $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, yields

$$\sum_{t=1}^{n} B_{2nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.28}$$

Similarly, it follows from (S.10), (S.16), Assumption 2 and the Hölder inequality that

$$E \sup_{\theta\in\Theta_n}\left|\sum_{t=1}^n B_{3nt}(\theta)\right| \leqslant E\sum_{t=1}^n \sup_{\theta\in\Theta_n}\left|[h_t^{-1}(\theta)-h_t^{-1}]\xi_{1nt}(\theta)\right| I\left(|\breve{e}_{t,\tau}| \leqslant \sup_{\theta\in\Theta_n}|\xi_{1nt}(\theta)|\right) = o(1),$$

and then

$$\sum_{t=1}^n B_{3nt}(\widetilde{\theta}_n) = o_p(1). \tag{S.29}$$

Finally, for $\sum_{t=1}^n B_{4nt}(\widetilde{\theta}_n)$, denote

$$B_{4nt}^*(\theta) = h_t^{-1}\int_0^{\xi_{1nt}(\theta)} \left[F(b_\tau - h_t^{-1}a_t + h_t^{-1}s) - F(b_\tau - h_t^{-1}a_t)\right] ds,$$

and we first show that

$$\sum_{t=1}^n B_{4nt}(\widetilde{\theta}_n) = \sum_{t=1}^n B_{4nt}^*(\widetilde{\theta}_n) + o_p(1). \tag{S.30}$$

For any $v\in\mathbb{R}^{p+q+1}$, let $\eta_t(v) = h_t^{-1}\int_0^{\xi_{1nt}(\theta_0+n^{-1/2}v)} I_t^*(s)ds$, and denote

$$S_n(v) = \sum_{t=1}^n \left[B_{4nt}(\theta_0+n^{-1/2}v) - B_{4nt}^*(\theta_0+n^{-1/2}v)\right] = \sum_{t=1}^n \left\{\eta_t(v) - E[\eta_t(v)|\mathcal{F}_{t-1}]\right\}.$$

For any fixed v such that $\|v\|\leqslant M$, by (S.16), Lemma A.1 and Assumption 2, we have

$$\begin{aligned} E\eta_t^2(v) &\leqslant E\left\{\frac{|\xi_{1nt}(\theta_0+n^{-1/2}v)|}{h_t^2}\int_0^{\xi_{1nt}(\theta_0+n^{-1/2}v)}[F(b_\tau - \frac{a_t}{h_t}+\frac{s}{h_t}) - F(b_\tau - \frac{a_t}{h_t})]ds\right\} \\ &\leqslant \frac{1}{2}\sup_{x\in\mathbb{R}} f(x) E|h_t^{-1}\xi_{1nt}(\theta_0+n^{-1/2}v)|^3 \leqslant n^{-3/2}C, \end{aligned} \tag{S.31}$$

implying that

$$ES_n^2(v) \leqslant \sum_{t=1}^n E\eta_t^2(v) = o(1). \tag{S.32}$$

Note that

$$h_t^{-1}\sup_{\|v_1-v_2\|\leqslant\delta}|\xi_{1nt}(\theta_0+n^{-1/2}v_1) - \xi_{1nt}(\theta_0+n^{-1/2}v_2)| \leqslant \frac{\delta}{\underline{w}\sqrt{n}}\sum_{j=1}^p |\beta_{\tau0}^{(j)}|\left\|\frac{1}{h_{t-j}}\frac{\partial h_{t-j}(\theta_0)}{\partial\theta}\right\|.$$

Then, for any $v_1, v_2\in\mathbb{R}^{p+q+1}$ such that $\|v_1\|,\|v_2\|\leqslant M$, in view of (S.10), (S.16), Lemma A.1 and Assumption 2, we have

$$\begin{aligned} &E\sup_{\|v_1-v_2\|\leqslant\delta}|\eta_t(v_1)-\eta_t(v_2)| \\ &= E\left[h_t^{-1}\sup_{\|v_1-v_2\|\leqslant\delta}\left|\int_{\xi_{1nt}(\theta_0+n^{-1/2}v_2)}^{\xi_{1nt}(\theta_0+n^{-1/2}v_1)} I_t^*(s)ds\right|\right] \\ &\leqslant E\left[h_t^{-1}\sup_{\|v_1-v_2\|\leqslant\delta}|\xi_{1nt}(\theta_0+n^{-1/2}v_1) - \xi_{1nt}(\theta_0+n^{-1/2}v_2)| I\left(|\breve{e}_{t,\tau}| \leqslant \sup_{\theta\in\Theta_n}|\xi_{1nt}(\theta)|\right)\right] \\ &\leqslant \frac{2\delta}{\underline{w}\sqrt{n}}\sup_{x\in\mathbb{R}} f(x) E\left(\sup_{\theta\in\Theta_n}\frac{|\xi_{1nt}(\theta)|}{h_t}\sum_{j=1}^p|\beta_{\tau0}^{(j)}|\left\|\frac{1}{h_{t-j}}\frac{\partial h_{t-j}(\theta_0)}{\partial\theta}\right\|\right) \leqslant n^{-1}\delta C, \end{aligned}$$

and hence

$$E \sup_{\|v_1-v_2\|\leqslant\delta} |S_n(v_1) - S_n(v_2)| \leqslant 2\sum_{t=1}^{n} E \sup_{\|v_1-v_2\|\leqslant\delta} |\eta_t(v_1) - \eta_t(v_2)| \leqslant 2\delta C,$$

which, together with (S.32) and the finite covering theorem, implies $\sup_{\|v\|\leqslant M} |S_n(v)| = o_p(1)$, and then (S.30) holds.

By elementary calculation and the Taylor expansion, we have

$$\begin{aligned}\sum_{t=1}^{n} B^*_{4nt}(\theta) &= \sum_{t=1}^{n} h_t^{-1} \int_0^{\xi_{1nt}(\theta)} f(b_\tau - h_t^{-1}a_t)h_t^{-1} s ds + R_{1n}(\theta) \\ &= \frac{1}{2} f(b_\tau) \sum_{t=1}^{n} h_t^{-2}\xi^2_{1nt}(\theta) + R_{2n}(\theta) + R_{1n}(\theta), \end{aligned} \tag{S.33}$$

where

$$R_{1n}(\theta) = \frac{1}{2}\sum_{t=1}^{n} h_t^{-3} \int_0^{\xi_{1nt}(\theta)} \dot{f}(b^*_{\tau,t}(s))s^2 ds,$$

with $b^*_{\tau,t}(s)$ lying between $b_\tau - h_t^{-1}a_t$ and $b_\tau - h_t^{-1}a_t + h_t^{-1}s$, and

$$R_{2n}(\theta) = \frac{1}{2}\sum_{t=1}^{n} h_t^{-2}\xi^2_{1nt}(\theta)[f(b_\tau - h_t^{-1}a_t) - f(b_\tau)].$$

Note that

$$\sup_{\theta\in\Theta_n} |R_{1n}(\theta)| \leqslant \frac{1}{6}\sup_{x\in\mathbb{R}} |\dot{f}(x)| \sum_{t=1}^{n} \sup_{\theta\in\Theta_n} \left|\frac{\xi_{1nt}(\theta)}{h_t}\right|^3,$$

and by Lemma S.1,

$$\sup_{\theta\in\Theta_n} |R_{2n}(\theta)| \leqslant \frac{1}{2} C \sup_{x\in\mathbb{R}} |\dot{f}(x)| \zeta \sum_{t=1}^{n} \rho^t \sup_{\theta\in\Theta_n} \left|\frac{\xi_{1nt}(\theta)}{h_t}\right|^2.$$

Then, by (S.16), Lemma A.1 and Assumption 2, we have

$$R_{1n}(\widetilde{\theta}_n) = o_p(1) \quad \text{and} \quad R_{2n}(\widetilde{\theta}_n) = o_p(1).$$

Hence, by (S.25), (S.27)-(S.30) and (S.33), together with the ergodic theorem, we have

$$\begin{aligned} L_{2n}(u) &= \frac{1}{2} f(b_\tau) \sum_{t=1}^{n} h_t^{-2}\xi^2_{1nt}(\widetilde{\theta}_n) + o_p(1) \\ &= \frac{1}{2} f(b_\tau) u'\Omega_2 u + b_\tau f(b_\tau) u'\Gamma_2 \sqrt{n}(\widetilde{\theta}_n - \theta_0) + T_{3n} + o_p(1), \end{aligned} \tag{S.34}$$

where

$$T_{3n} = \frac{1}{2} f(b_\tau)(\widetilde{\theta}_n - \theta_0)' \sum_{t=1}^{n}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p} \beta^{(j_1)}_{\tau 0}\beta^{(j_2)}_{\tau 0} \frac{1}{h_t^2} \frac{\partial h_{t-j_1}(\theta_0)}{\partial\theta} \frac{\partial h_{t-j_2}(\theta_0)}{\partial\theta'} (\widetilde{\theta}_n - \theta_0).$$

Combining (S.6), (S.24) and (S.34) yields that

$$L_n(\theta_{\tau 0}+n^{-1/2}u)-\breve{L}_n(\theta_{\tau 0}) = -u'\left[T_{1n}-b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\widetilde{\theta}_n-\theta_0)\right]+\frac{1}{2}f(b_\tau)u'\Omega_2 u \\ -T_{2n}+T_{3n}+o_p(1),$$

where

$$\sqrt{n}(\widetilde{\theta}_n-\theta_0) = -\frac{J^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}+o_p(1); \tag{S.35}$$

see Francq and Zakoian (2004). By the central limit theorem and Corollary 2 in Knight (1998), together with the convexity of $L_n(\cdot)$, we have

$$\sqrt{n}(\widehat{\theta}_{\tau n}-\theta_{\tau 0}) = \frac{\Omega_2^{-1}}{f(b_\tau)}T_{1n}-b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\widetilde{\theta}_n-\theta_0)+o_p(1)\to_d N(0,\Sigma_1), \tag{S.36}$$

where $T_{1n}=n^{-1/2}\sum_{t=1}^n\psi_\tau(\varepsilon_t-b_\tau)z_t/h_t$; see also Lemma 2.2 of Davis et al. (1992). The proof is complete. □

Proof of Theorem 2. Let $L_n^*(\theta)=\sum_{t=1}^n\omega_t\widetilde{h}_t^{-1}\rho_\tau(y_t-\theta'\widetilde{z}_t^*)$ and $\breve{L}_n^*(\theta)=\sum_{t=1}^n\omega_t\widetilde{h}_t^{-1}\rho_\tau(y_t-\theta'\breve{z}_t)$. For any fixed $u\in\mathbb{R}^{p+q+1}$, similar to (S.6), it holds that

$$L_n^*(\theta_{\tau 0}+n^{-1/2}u)-\breve{L}_n^*(\theta_{\tau 0}) = -L_{1n}^*(u)+L_{2n}^*(u), \tag{S.37}$$

where

$$L_{1n}^*(u) = \sum_{t=1}^n\omega_t\psi_\tau(\breve{e}_{t,\tau})\widetilde{h}_t^{-1}\left[(\theta_{\tau 0}+n^{-1/2}u)'\widetilde{z}_t^*-\theta_{\tau 0}'\breve{z}_t\right],$$
$$L_{2n}^*(u) = \sum_{t=1}^n\omega_t\widetilde{h}_t^{-1}\int_0^{(\theta_{\tau 0}+n^{-1/2}u)'\widetilde{z}_t^*-\theta_{\tau 0}'\breve{z}_t}\left[I(\breve{e}_{t,\tau}\leqslant s)-I(\breve{e}_{t,\tau}\leqslant 0)\right]ds,$$

and

$$(\theta_{\tau 0}+n^{-1/2}u)'\widetilde{z}_t^*-\theta_{\tau 0}'\breve{z}_t=\xi_{1nt}(\widetilde{\theta}_n^*)+\xi_{2nt}(\widetilde{\theta}_n^*)+\xi_{3nt}(\widetilde{\theta}_n^*).$$

From the proof of Theorem 1, we have $\widetilde{J}=J+o_p(1)$, which together with (3.4) implies

$$\sqrt{n}(\widetilde{\theta}_n^*-\widetilde{\theta}_n) = -\frac{J^{-1}}{\sqrt{n}}\sum_{t=1}^n(\omega_t-1)\left(1-\frac{|y_t|}{h_t}\right)\frac{1}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}+o_p^*(1), \tag{S.38}$$

and

$$\sqrt{n}(\widetilde{\theta}_n^*-\theta_0)=\sqrt{n}(\widetilde{\theta}_n^*-\widetilde{\theta}_n)+\sqrt{n}(\widetilde{\theta}_n-\theta_0)=O_p^*(1). \tag{S.39}$$

Without any confusion, we redefine the functions A_{int} with $1 \leqslant i \leqslant 4$ from the proof of Theorem 1 as follows,

$$A_{1nt}(\theta_1,\theta_2) = \psi_\tau(\breve{e}_{t,\tau})\widetilde{h}_t^{-1}(\theta_1)\xi_{3nt}(\theta_2) + \psi_\tau(\breve{e}_{t,\tau})[\widetilde{h}_t^{-1}(\theta_1) - h_t^{-1}(\theta_1)][\xi_{1nt}(\theta_2) + \xi_{2nt}(\theta_2)],$$
$$A_{2nt}(\theta_1,\theta_2) = [\psi_\tau(\breve{e}_{t,\tau}) - \psi_\tau(\varepsilon_t - b_\tau)]h_t^{-1}(\theta_1)[\xi_{1nt}(\theta_2) + \xi_{2nt}(\theta_2)],$$
$$A_{3nt}(\theta_1,\theta_2) = \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}(\theta_1)\xi_{2nt}(\theta_2), \text{ and } A_{4nt}(\theta_1,\theta_2) = \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}(\theta_1)\xi_{1nt}(\theta_2),$$

as well as B_{int} with $1 \leqslant i \leqslant 3$ as follows,

$$\begin{aligned}
B_{1nt}(\theta_1,\theta_2) &= \widetilde{h}_t^{-1}(\theta_1)\int_{\xi_{1nt}(\theta_2)+\xi_{2nt}(\theta_2)}^{\xi_{1nt}(\theta_2)+\xi_{2nt}(\theta_2)+\xi_{3nt}(\theta_2)} I_t^*(s)ds \\
&\quad + [\widetilde{h}_t^{-1}(\theta_1) - h_t^{-1}(\theta_1)]\int_0^{\xi_{1nt}(\theta_2)+\xi_{2nt}(\theta_2)} I_t^*(s)ds, \\
B_{2nt}(\theta_1,\theta_2) &= h_t^{-1}(\theta_1)\int_{\xi_{1nt}(\theta_2)}^{\xi_{1nt}(\theta_2)+\xi_{2nt}(\theta_2)} I_t^*(s)ds, \quad \text{and} \\
B_{3nt}(\theta_1,\theta_2) &= [h_t^{-1}(\theta_1) - h_t^{-1}]\int_0^{\xi_{1nt}(\theta_2)} I_t^*(s)ds,
\end{aligned}$$

while the definition of $B_{4nt}(\cdot)$ is the same as in the proof of Theorem 1.

By methods similar to (S.13), (S.18), (S.22) and (S.23) respectively, together with Assumption 2, Lemma S.1, (S.8), (S.9) and (S.38), we can show that

$$\sum_{t=1}^n \omega_t A_{int}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) = o_p^*(1), \quad 1 \leqslant i \leqslant 3,$$

and

$$\sum_{t=1}^n \omega_t A_{4nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) = \sum_{t=1}^n \omega_t \psi_\tau(\varepsilon_t - b_\tau)h_t^{-1}\xi_{1nt}(\widetilde{\theta}_n^*) + o_p^*(1) = u'T_{1n}^* + T_{2n}^* + o_p^*(1),$$

where $T_{1n}^* = n^{-1/2}\sum_{t=1}^n \omega_t\psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$ and

$$T_{2n}^* = \sqrt{n}(\widetilde{\theta}_n^* - \theta_0)'\frac{1}{\sqrt{n}}\sum_{t=1}^n \omega_t\psi_\tau(\varepsilon_t - b_\tau)\sum_{j=1}^p \frac{\beta_{\tau 0}^{(j)}}{h_t}\frac{\partial h_{t-j}(\theta_0)}{\partial \theta},$$

where $\beta_{\tau 0}^{(j)} = b_\tau\beta_{0j}$, $j = 1, \ldots, p$, is defined as in the proof of Theorem 1. As a result,

$$\begin{aligned}
L_{1n}^*(u) &= \sum_{t=1}^n \omega_t A_{1nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n \omega_t A_{2nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n \omega_t A_{3nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n \omega_t A_{4nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) \\
&= u'T_{1n}^* + T_{2n}^* + o_p^*(1). \qquad \text{(S.40)}
\end{aligned}$$

Moreover, by methods similar to (S.27)-(S.29), we can verify that

$$\sum_{t=1}^n (\omega_t - 1)B_{int}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) = o_p^*(1), \ 1 \leqslant i \leqslant 3, \text{ and } \sum_{t=1}^n (\omega_t - 1)B_{4nt}(\widetilde{\theta}_n^*) = o_p^*(1),$$

which implies

$$L_{2n}^*(u) = \sum_{t=1}^n B_{1nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n B_{2nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n B_{3nt}(\widetilde{\theta}_n, \widetilde{\theta}_n^*) + \sum_{t=1}^n B_{4nt}(\widetilde{\theta}_n^*) + o_p^*(1),$$

and hence, similar to the proof of (S.34), it can be further verified that

$$L_{2n}^*(u) = \frac{1}{2}f(b_\tau)u'\Omega_2 u + b_\tau f(b_\tau)u'\Gamma_2\sqrt{n}(\widetilde{\theta}_n^* - \theta_0) + T_{3n}^* + o_p^*(1), \tag{S.41}$$

where

$$T_{3n}^* = \frac{1}{2}f(b_\tau)(\widetilde{\theta}_n^* - \theta_0)'\sum_{t=1}^n\sum_{j_1=1}^p\sum_{j_2=1}^p \beta_{\tau 0}^{(j_1)}\beta_{\tau 0}^{(j_2)}\frac{1}{h_t^2}\frac{\partial h_{t-j_1}(\theta_0)}{\partial\theta}\frac{\partial h_{t-j_2}(\theta_0)}{\partial\theta'}(\widetilde{\theta}_n^* - \theta_0).$$

Therefore, combining (S.37), (S.40) and (S.41), we have

$$\begin{aligned} L_n^*(\theta_{\tau 0} + n^{-1/2}u) - \breve{L}_n^*(\theta_{\tau 0}) = & - u'\left[T_{1n}^* - b_\tau f(b_\tau)\Gamma_2\sqrt{n}(\widetilde{\theta}_n^* - \theta_0)\right] + \frac{1}{2}f(b_\tau)u'\Omega_2 u \\ & - T_{2n}^* + T_{3n}^* + o_p^*(1), \end{aligned}$$

where $T_{1n}^* = n^{-1/2}\sum_{t=1}^n \omega_t\psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$.

Denote $X_t = n^{-1/2}(\omega_t - 1)\psi_\tau(\varepsilon_t - b_\tau)z_t/h_t$, and then $T_{1n}^* - T_{1n} = \sum_{t=1}^n X_t$. For any constant vector $c \in \mathbb{R}^{p+q+1}$, let $\mu_t = E^*(c'X_t)$ and $\sigma_n^2 = \sum_{t=1}^n E^*(c'X_tX_t'c)$. Then, $\mu_t = 0$, and by (S.15) we have

$$\begin{aligned} \left(\sum_{t=1}^n E^*|c'X_t - \mu_t|^{2+\delta}\right)^{\frac{1}{2+\delta}} & = \frac{1}{\sqrt{n}}\left[\sum_{t=1}^n\left|\psi_\tau(\varepsilon_t - b_\tau)\frac{c'z_t}{h_t}\right|^{2+\delta}\right]^{\frac{1}{2+\delta}}(E^*|\omega_t - 1|^{2+\delta})^{\frac{1}{2+\delta}} \\ & = o_p(1), \end{aligned}$$

as long as $0 < \delta \leqslant \kappa_0$, since $E^*|\omega_t|^{2+\kappa_0} < \infty$ from the assumptions of this theorem. Moreover, by the ergodic theorem, $\sigma_n^2 = c'n^{-1}\sum_{t=1}^n[\psi_\tau(\varepsilon_t - b_\tau)]^2h_t^{-2}z_tz_t'c = \tau(1-\tau)c'\Omega_2 c + o_p(1)$. Thus, we can show that the Liapounov's condition, $\sum_{t=1}^n E^*|c'X_t - \mu_t|^{2+\delta} = o_p(\sigma_n^{2+\delta})$, holds for $0 < \delta \leqslant \kappa_0$. This, together with the Cramér-Wold device and the Lindeberg's central limit theorem, implies that conditional on $\mathcal{F}_n$,

$$T_{1n}^* - T_{1n} = \sum_{t=1}^n X_t \to_d N(0, \tau(1-\tau)\Omega_2)$$

in probability as $n \to \infty$.

Since $L_n^*(\cdot)$ is convex, by Corollary 2 of Knight (1998), it holds that

$$\sqrt{n}(\widehat{\theta}_{\tau n}^* - \theta_{\tau 0}) = \frac{\Omega_2^{-1}}{f(b_\tau)}T_{1n}^* - b_\tau\Omega_2^{-1}\Gamma_2\sqrt{n}(\widetilde{\theta}_n^* - \theta_0) + o_p^*(1), \tag{S.42}$$

which, in conjunction with (S.36), yields the Bahadur representation of the corrected bootstrap estimator $\widehat{\theta}^*_{\tau n}$,

$$\sqrt{n}(\widehat{\theta}^*_{\tau n} - \widehat{\theta}_{\tau n}) = \frac{\Omega_2^{-1}}{f(b_\tau)}(T^*_{1n} - T_{1n}) + \frac{b_\tau \Omega_2^{-1}\Gamma_2 J^{-1}}{\sqrt{n}}\sum_{t=1}^{n}(\omega_t - 1)\frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} + o_p^*(1).$$

Denote $X_t^\dagger = n^{-1/2}(\omega_t - 1)d_t$, with $d_t = (\psi_\tau(\varepsilon_t - b_\tau)z_t'/h_t, (1-|\varepsilon_t|)h_t^{-1}\partial h_t(\theta_0)/\partial\theta')'$. Note that by (S.15) and $E|\varepsilon_t|^{2+\nu_0} < \infty$ for $\nu_0 > 0$, we have $E|d_t|^{2+\nu_0} < \infty$. Then, for $0 < \delta \leqslant \min(\kappa_0, \nu_0)$, we can similarly verify the Liapounov's condition, $\sum_{t=1}^n E^*|c'X_t^\dagger - \mu_t^\dagger|^{2+\delta} = o_p(\sigma_n^{\dagger 2+\delta})$, where $\mu_t^\dagger = E^*(c'X_t^\dagger)$ and $\sigma_n^{\dagger 2} = \sum_{t=1}^n E^*(c'X_t^\dagger X_t^{\dagger\prime} c)$. Applying the Lindeberg's central limit theorem and the Cramér-Wold device, we accomplish the proof of the theorem. □

Proof of Theorem 3. Observe that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\psi_\tau(\widehat{\varepsilon}_{t,\tau})|\widehat{\varepsilon}_{t-k,\tau}| \\ &\quad= \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}| + \sum_{t=k+1}^{n}\mathcal{E}_{1nt} + \sum_{t=k+1}^{n}\mathcal{E}_{2nt} + \sum_{t=k+1}^{n}\mathcal{E}_{3nt},\end{aligned} \tag{S.43}$$

where

$$\mathcal{E}_{1nt} = n^{-1/2}[\psi_\tau(\widehat{\varepsilon}_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})]|\varepsilon_{t-k,\tau}|, \quad \mathcal{E}_{2nt} = n^{-1/2}\psi_\tau(\varepsilon_{t,\tau})(|\widehat{\varepsilon}_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|), \quad \text{and}$$
$$\mathcal{E}_{3nt} = n^{-1/2}[\psi_\tau(\widehat{\varepsilon}_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})](|\widehat{\varepsilon}_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|).$$

To derive the asymptotic result for the quantity on the left-hand side of (S.43), we shall begin by proving that

$$\sum_{t=k+1}^{n}\mathcal{E}_{1nt} = -f(b_\tau)\left[d_{1k}'\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) + b_\tau d_{2k}'\sqrt{n}(\widetilde{\theta}_n - \theta_0)\right] + o_p(1), \tag{S.44}$$

where $d_{1k} = E(h_t^{-1}|\varepsilon_{t-k,\tau}|z_t)$ and $d_{2k} = E(h_t^{-1}|\varepsilon_{t-k,\tau}|\sum_{j=1}^p \beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta)$. For any $u, v \in \mathbb{R}^{p+q+1}$, define

$$\widetilde{b}_t(u,v) = (\theta_{\tau 0} + n^{-1/2}u)'\widetilde{z}_t(\theta_0 + n^{-1/2}v)h_t^{-1}.$$

Since $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = O_p(1)$, $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, and

$$\sum_{t=k+1}^{n}\mathcal{E}_{1nt} = \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\left[I(\varepsilon_t < b_\tau) - I(\varepsilon_t < \widehat{\theta}_{\tau n}'\widetilde{z}_t h_t^{-1})\right]|\varepsilon_{t-k,\tau}|,$$

to prove (S.44), it suffices to show that for any $M > 0$,

$$\sup_{\|u\|,\|v\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\phi_t(u,v) + f(b_\tau)\left(d_{1k}'u + b_\tau d_{2k}'v\right)\right| = o_p(1), \tag{S.45}$$

where $\phi_t(u,v) = \{I(\varepsilon_t < b_\tau) - I[\varepsilon_t < \widetilde{b}_t(u,v)]\}|\varepsilon_{t-k,\tau}|$.

Let $S_n(u,v) = n^{-1/2}\sum_{t=k+1}^n \{\phi_t(u,v) - E[\phi_t(u,v)|\mathcal{F}_{t-1}]\}$, and we shall first show that

$$\sup_{\|u\|,\|v\|\leqslant M} |S_n(u,v)| = o_p(1). \tag{S.46}$$

For any $u, v \in \mathbb{R}^{p+q+1}$, define

$$b_t(u,v) = (\theta_{\tau 0} + n^{-1/2}u)' z_t(\theta_0 + n^{-1/2}v) h_t^{-1}.$$

Note that for any $u_i, v_i \in \mathbb{R}^{p+q+1}$, $i = 1, 2$, since

$$\begin{aligned} & b_t(u_1,v_1) - b_t(u_2,v_2) \\ & \quad = \sum_{j=1}^p \frac{\beta_{\tau 0}^{(j)} + n^{-1/2}u_1^{(j)}}{h_t}[h_{t-j}(\theta_0 + n^{-1/2}v_1) - h_{t-j}(\theta_0 + n^{-1/2}v_2)] \\ & \qquad + \frac{1}{\sqrt{n}}\sum_{j=1}^p \frac{u_1^{(j)} - u_2^{(j)}}{h_t}[h_{t-j}(\theta_0 + n^{-1/2}v_2) - h_{t-j}] + \frac{h_t^{-1} z_t'(u_1 - u_2)}{\sqrt{n}}, \end{aligned}$$

by the Taylor expansion and (S.15), where $\beta_{\tau 0}^{(j)} = b_\tau \beta_{0j}$ for $j = 1, \ldots, p$, we can readily show that if $\|u_i\|, \|v_i\| \leqslant M$, then

$$\begin{aligned} & |b_t(u_1,v_1) - b_t(u_2,v_2)| \\ & \leqslant \frac{C(M)}{\sqrt{n}}\left[\left(\|v_1 - v_2\| + \frac{\|u_1 - u_2\|}{\sqrt{n}}\right)\sum_{j=1}^p \sup_{\theta\in\Theta_n}\left\|\frac{1}{h_{t-j}}\frac{\partial h_{t-j}(\theta)}{\partial\theta}\right\| + \|u_1 - u_2\|\right]. \end{aligned} \tag{S.47}$$

For any $u, v \in \mathbb{R}^{p+q+1}$ such that $\|u\|, \|v\| \leqslant M$, by the Hölder inequality and the fact that $E|\varepsilon_t|^{2+\nu_0} < \infty$ for $\nu_0 > 0$, we have

$$\begin{aligned} \sum_{t=k+1}^n E\phi_t^2(u,v) & \leqslant \sum_{t=k+1}^n \left\{E\left|I(\varepsilon_t < b_\tau) - I[\varepsilon_t < \widetilde{b}_t(u,v)]\right|\right\}^{\frac{\nu_0}{2+\nu_0}} \left(E|\varepsilon_{t-k,\tau}|^{2+\nu_0}\right)^{\frac{2}{2+\nu_0}} \\ & = C\sum_{t=k+1}^n \left[E\left|F(\widetilde{b}_t(u,v)) - F(b_\tau)\right|\right]^{\frac{\nu_0}{2+\nu_0}} \\ & \leqslant C\Bigg\{\sum_{t=k+1}^n \left[E\left|F(\widetilde{b}_t(u,v)) - F(b_t(u,v))\right|\right]^{\frac{\nu_0}{2+\nu_0}} \\ & \qquad + \sum_{t=k+1}^n [E|F(b_t(u,v)) - F(b_\tau)|]^{\frac{\nu_0}{2+\nu_0}}\Bigg\}, \end{aligned} \tag{S.48}$$

where the last inequality follows from the fact that $(x+y)^a \leqslant x^a + y^a$ for any $x, y \geqslant 0$ and $0 < a < 1$. Note that by Lemma S.1, we have

$$\begin{aligned} \sup_{\|u\|,\|v\|\leqslant M} |\widetilde{b}_t(u,v) - b_t(u,v)| & \leqslant \sum_{j=1}^p \frac{|\beta_{\tau 0}^{(j)}| + n^{-1/2}M}{\underline{w}} \sup_{\theta\in\Theta} |\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)| \\ & \leqslant C(M)\rho^t\zeta. \end{aligned} \tag{S.49}$$

Then, by Assumption 2 and a method similar to that for (S.14), we can show that

$$E\left|F(\widetilde{b}_t(u,v)) - F(b_t(u,v))\right| \leqslant \rho^{t/2} + C(M)\rho^{\delta_0 t/2}. \tag{S.50}$$

Moreover, since $b_\tau = b_t(0,0)$, it follows from (S.47), Lemma A.1 and Assumption 2 that

$$E|F(b_t(u,v)) - F(b_\tau)| \leqslant \sup_{x\in\mathbb{R}} f(x)E|b_t(u,v) - b_\tau| \leqslant n^{-1/2}C(M). \tag{S.51}$$

In view of (S.48), (S.50) and (S.51), for any $u, v \in \mathbb{R}^{p+q+1}$ with $\|u\|, \|v\| \leqslant M$,

$$ES_n^2(u,v) \leqslant \frac{1}{n}\sum_{t=k+1}^{n} E\phi_t^2(u,v) = o(1). \tag{S.52}$$

For any $\delta > 0$, let $U(\delta)$ be the set of all four-tuples (u_1, u_2, v_1, v_2) of column vectors in $\mathbb{R}^{p+q+1}$ such that $\|u_i\|, \|v_i\| \leqslant M$, $i = 1, 2$, and $\|u_1 - u_2\|, \|v_1 - v_2\| \leqslant \delta$, and denote by υ an element of $U(\delta)$. Moreover, for simplicity, denote $\widetilde{b}_{ti} = \widetilde{b}_t(u_i, v_i)$ and $b_{ti} = b_t(u_i, v_i)$ for $i = 1, 2$. Let $\widetilde{\Delta}_t = \sup_{\upsilon\in U(\delta)} |\widetilde{b}_{t1} - \widetilde{b}_{t2}|$ and $\Delta_t = \sup_{\upsilon\in U(\delta)} |b_{t1} - b_{t2}|$. Notice that

$$\begin{aligned}\sup_{\upsilon\in U(\delta)} |\phi_t(u_1,v_1) - \phi_t(u_2,v_2)| &\leqslant \sup_{\upsilon\in U(\delta)} \left|I(\varepsilon_t < \widetilde{b}_{t2}) - I(\varepsilon_t < \widetilde{b}_{t1})\right||\varepsilon_{t-k,\tau}| \\ &\leqslant I\big(|\varepsilon_t - \widetilde{b}_{t2}| < \widetilde{\Delta}_t\big)|\varepsilon_{t-k,\tau}|.\end{aligned}$$

Then, applying the Hölder inequality, together with $E|\varepsilon_t|^{2+\nu_0} < \infty$ for $\nu_0 > 0$ and the fact that $(x+y)^a \leqslant x^a + y^a$ for any $x, y \geqslant 0$ and $0 < a < 1$, we have

$$\begin{aligned}&E\sup_{\upsilon\in U(\delta)} |\phi_t(u_1,v_1) - \phi_t(u_2,v_2)| \\ &\quad\leqslant \Big[E\big|F\big(\widetilde{b}_{t2} + \widetilde{\Delta}_t\big) - F\big(\widetilde{b}_{t2} - \widetilde{\Delta}_t\big)\big|\Big]^{1/2}(E\varepsilon_{t-k,\tau}^2)^{1/2} \\ &\quad\leqslant C\bigg\{\Big[E\big|F\big(\widetilde{b}_{t2} + \widetilde{\Delta}_t\big) - F\big(\widetilde{b}_{t2} + \Delta_t\big)\big|\Big]^{1/2} + \Big[E\big|F\big(\widetilde{b}_{t2} - \widetilde{\Delta}_t\big) - F\big(\widetilde{b}_{t2} - \Delta_t\big)\big|\Big]^{1/2} \\ &\qquad\quad + \Big[E\big|F\big(\widetilde{b}_{t2} + \Delta_t\big) - F\big(\widetilde{b}_{t2} - \Delta_t\big)\big|\Big]^{1/2}\bigg\}.\end{aligned} \tag{S.53}$$

Since $|\widetilde{\Delta}_t - \Delta_t| \leqslant \sup_{\upsilon\in U(\delta)} \left|(\widetilde{b}_{t1} - \widetilde{b}_{t2}) - (b_{t1} - b_{t2})\right| \leqslant 2\sup_{\|u\|,\|v\|\leqslant M} |\widetilde{b}_t(u,v) - b_t(u,v)|$, by (S.49) and a method similar to that for (S.14), we can verify that

$$E\big|F\big(\widetilde{b}_{t2} \pm \widetilde{\Delta}_t\big) - F\big(\widetilde{b}_{t2} \pm \Delta_t\big)\big| \leqslant \rho^{t/2} + C(M)\rho^{\delta_0 t/2}. \tag{S.54}$$

Furthermore, it follows from Assumption 2, (S.47) and Lemma A.1 that

$$E\big|F\big(\widetilde{b}_{t2} + \Delta_t\big) - F\big(\widetilde{b}_{t2} - \Delta_t\big)\big| \leqslant 2\sup_{x\in\mathbb{R}} f(x)E(\Delta_t) \leqslant n^{-1/2}\delta C(M). \tag{S.55}$$

As a result of (S.53)-(S.55), we have

$$E \sup_{v\in U(\delta)} |S_n(u_1,v_1) - S_n(u_2,v_2)| \leqslant \frac{2}{\sqrt{n}} \sum_{t=k+1}^{n} E \sup_{v\in U(\delta)} |\phi_t(u_1,v_1) - \phi_t(u_2,v_2)| \leqslant \delta C(M),$$

which, together with (S.52) and the finite covering theorem, implies (S.46).

Since $E[\phi_t(u,v)|\mathcal{F}_{t-1}] = [F(b_\tau) - F(\widetilde{b}_t(u,v))]|\varepsilon_{t-k,\tau}|$, to prove (S.45), it remains to show that

$$\sup_{\|u\|,\|v\|\leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \big[F(b_\tau) - F(\widetilde{b}_t(u,v))\big]|\varepsilon_{t-k,\tau}| + f(b_\tau)\left(d_{1k}'u + b_\tau d_{2k}'v\right) \right| = o_p(1). \quad \text{(S.56)}$$

By (S.49), Assumption 2 and a method similar to that for (S.14), we can show that

$$E\bigg(\sup_{\|u\|,\|v\|\leqslant M} \Big| F(\widetilde{b}_t(u,v)) - F(b_t(u,v)) \Big| \bigg)^2 \leqslant \rho^t + C(M)\rho^{\delta_0 t/2},$$

which, in conjunction with the Hölder inequality and $E|\varepsilon_t|^{2+\nu_0} < \infty$ for $\nu_0 > 0$, yields

$$\begin{aligned} &E \sup_{\|u\|,\|v\|\leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \big[F(\widetilde{b}_t(u,v)) - F(b_t(u,v))\big]|\varepsilon_{t-k,\tau}| \right| \\ &\quad \leqslant \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \left[E\bigg(\sup_{\|u\|,\|v\|\leqslant M} \Big| F(b_t(u,v)) - F(\widetilde{b}_t(u,v)) \Big| \bigg)^2 \right]^{1/2} (E\varepsilon_{t-k,\tau}^2)^{1/2} = o(1), \end{aligned}$$

and hence,

$$\sup_{\|u\|,\|v\|\leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \big[F(b_t(u,v)) - F(\widetilde{b}_t(u,v))\big]|\varepsilon_{t-k,\tau}| \right| = o_p(1). \quad \text{(S.57)}$$

Note that by the Taylor expansion,

$$b_\tau - b_t(u,v) = -\frac{h_t^{-1} z_t' u}{\sqrt{n}} - \frac{v'}{\sqrt{n}} \sum_{j=1}^{p} \frac{\beta_{\tau 0}^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta} - R_t(u,v),$$

where

$$R_t(u,v) = \frac{v'}{n} \sum_{j=1}^{p} \frac{u^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta} + \frac{v'}{2n} \sum_{j=1}^{p} \frac{\beta_{\tau 0}^{(j)} + n^{-1/2} u^{(j)}}{h_t} \frac{\partial^2 h_{t-j}(\theta^*)}{\partial \theta \partial \theta'} v,$$

with θ^* between θ_0 and $\theta_0 + n^{-1/2}v$. Then, by (S.47), Assumption 2, Lemma A.1 and the ergodic theorem, we can show that

$$\begin{aligned} &\sup_{\|u\|,\|v\|\leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \big[F(b_\tau) - F(b_t(u,v))\big]|\varepsilon_{t-k,\tau}| + f(b_\tau)\left(d_{1k}'u + b_\tau d_{2k}'v\right) \right| \\ &\quad \leqslant f(b_\tau) \sup_{\|u\|,\|v\|\leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} [b_\tau - b_t(u,v)]|\varepsilon_{t-k,\tau}| + d_{1k}'u + b_\tau d_{2k}'v \right| \\ &\qquad + \frac{1}{2} \sup_{x\in\mathbb{R}} |\dot{f}(x)| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \sup_{\|u\|,\|v\|\leqslant M} |b_\tau - b_t(u,v)|^2 |\varepsilon_{t-k,\tau}| \\ &\quad = o_p(1). \end{aligned}$$

This together with (S.57) implies (S.56), and therefore, (S.44) holds.

Next, we consider $\sum_{t=k+1}^{n} \mathcal{E}_{2nt}$. Observe that

$$\varepsilon_{t,\tau} - \widehat{\varepsilon}_{t,\tau} = \zeta_{1nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) + \zeta_{2nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n),$$

where

$$\zeta_{1nt}(\theta_\tau, \theta) = \frac{y_t - \theta_{\tau 0}' z_t}{h_t} - \frac{y_t - \theta_\tau' z_t(\theta)}{h_t(\theta)} \quad \text{and} \quad \zeta_{2nt}(\theta_\tau, \theta) = \frac{y_t - \theta_\tau' z_t(\theta)}{h_t(\theta)} - \frac{y_t - \theta_\tau' \widetilde{z}_t(\theta)}{\widetilde{h}_t(\theta)}.$$

Then, similar to the decompositions in (S.6), (S.11) and (S.25), by using the identity in (S.5), it can be verified that

$$\sum_{t=k+1}^{n} \mathcal{E}_{2nt} = \sum_{t=1}^{n-k} Z_{1nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) + \sum_{t=1}^{n-k} Z_{2nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) + \sum_{t=1}^{n-k} Z_{3nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n), \tag{S.58}$$

where

$$\begin{aligned}
Z_{1nt}(\theta_\tau, \theta) &= \frac{\psi_\tau(\varepsilon_{t+k,\tau})}{\sqrt{n}} \left\{ -\zeta_{2nt}(\theta_\tau, \theta)[1 - 2I(\varepsilon_t < b_\tau)] + 2\int_{\zeta_{1nt}(\theta_\tau,\theta)}^{\zeta_{1nt}(\theta_\tau,\theta)+\zeta_{2nt}(\theta_\tau,\theta)} I_t(s)ds \right\}, \\
Z_{2nt}(\theta_\tau, \theta) &= -\frac{\psi_\tau(\varepsilon_{t+k,\tau})}{\sqrt{n}} \zeta_{1nt}(\theta_\tau, \theta)[1 - 2I(\varepsilon_t < b_\tau)], \quad \text{and} \\
Z_{3nt}(\theta_\tau, \theta) &= \frac{2\psi_\tau(\varepsilon_{t+k,\tau})}{\sqrt{n}} \int_0^{\zeta_{1nt}(\theta_\tau,\theta)} I_t(s)ds,
\end{aligned}$$

with $I_t(s) = I(\varepsilon_{t,\tau} \leqslant s) - I(\varepsilon_{t,\tau} \leqslant 0)$. For any $M > 0$, let $\Theta_{\tau n} = \Theta_{\tau n}(M) = \{\theta_\tau : \|\theta_\tau - \theta_{\tau 0}\| \leqslant n^{-1/2}M, \theta_\tau/b_\tau \in \Theta\}$. Note that $\zeta_{2nt}(\theta_\tau, \theta) = \widetilde{h}_t^{-1}(\theta)\theta_\tau'[\widetilde{z}_t(\theta) - z_t(\theta)] + [h_t^{-1}(\theta) - \widetilde{h}_t^{-1}(\theta)][y_t - \theta_\tau' z_t(\theta)]$. Then, similar to (S.9), (S.12) and (S.26), by Lemma S.1, it can be shown that

$$\begin{aligned}
\sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{2nt}(\theta_\tau, \theta)| &\leqslant \frac{1}{\underline{w}} \sum_{j=1}^{p} \sup_{\theta_\tau \in \Theta_{\tau n}} |\beta_\tau^{(j)}| \sup_{\theta \in \Theta} |\widetilde{h}_{t-j}(\theta) - h_{t-j}(\theta)| \\
&\quad + \frac{1}{\underline{w}^2} \sup_{\theta \in \Theta} |\widetilde{h}_t(\theta) - h_t(\theta)| \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |y_t - \theta_\tau' z_t(\theta)| \\
&\leqslant C(M)\rho^t \zeta \left[1 + \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |y_t - \theta_\tau' z_t(\theta)| \right].
\end{aligned}$$

Consequently, it follows from Lemma A.1 that

$$\sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} \left| \sum_{t=1}^{n-k} Z_{1nt}(\theta_\tau, \theta) \right| \leqslant \frac{3}{\sqrt{n}} \sum_{t=1}^{n-k} \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{2nt}(\theta_\tau, \theta)| = o_p(1),$$

which, together with $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = O_p(1)$ and $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, yields

$$\sum_{t=1}^{n-k} Z_{1nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) = o_p(1). \tag{S.59}$$

Applying the second-order Taylor expansion to $h_t^{-1}(\theta)$, and the first and second-order Taylor expansions to $\theta_\tau' z_t(\theta)$ respectively, similar to (S.19), it can be verified that

$$\zeta_{1nt}(\theta_\tau, \theta) = \zeta_{3nt}(\theta_\tau, \theta) + \zeta_{4nt}(\theta_\tau, \theta), \tag{S.60}$$

where

$$\begin{aligned}
\zeta_{3nt}(\theta_\tau, \theta) =& (\theta_\tau - \theta_{\tau 0})' \frac{z_t}{h_t} + (\theta - \theta_0)' \sum_{j=1}^p \frac{\beta_{\tau 0}^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_0)}{\partial \theta} + (\theta - \theta_0)' \frac{\varepsilon_t - b_\tau}{h_t} \frac{\partial h_t(\theta_0)}{\partial \theta}, \\
\zeta_{4nt}(\theta_\tau, \theta) =& (\theta - \theta_0)' \sum_{j=1}^p \frac{\beta_\tau^{(j)} - \beta_{\tau 0}^{(j)}}{h_t} \frac{\partial h_{t-j}(\theta_2^*)}{\partial \theta} + \frac{(\theta - \theta_0)'}{2} \sum_{j=1}^p \frac{\beta_{\tau 2}^{*(j)}}{h_t} \frac{\partial^2 h_{t-j}(\theta_2^*)}{\partial \theta \partial \theta'} (\theta - \theta_0) \\
& - \frac{(\theta - \theta_0)'}{h_t} \frac{\partial h_t(\theta_0)}{\partial \theta} \left[\frac{z_t'(\theta_1^*)}{h_t} (\theta_\tau - \theta_{\tau 0}) + \sum_{j=1}^p \frac{\beta_{\tau 1}^{*(j)}}{h_t} \frac{\partial h_{t-j}(\theta_1^*)}{\partial \theta'} (\theta - \theta_0) \right] \\
& - \frac{y_t - \theta_\tau' z_t(\theta)}{h_t(\theta_3^*)} \frac{(\theta - \theta_0)'}{2} \left[\frac{2}{h_t^2(\theta_3^*)} \frac{\partial h_t(\theta_3^*)}{\partial \theta} \frac{\partial h_t(\theta_3^*)}{\partial \theta'} - \frac{1}{h_t(\theta_3^*)} \frac{\partial^2 h_t(\theta_3^*)}{\partial \theta \partial \theta'} \right] (\theta - \theta_0),
\end{aligned}$$

with θ_1^*, θ_2^* and θ_3^* all lying between θ_0 and θ, and $\beta_{\tau 1}^{*(j)}$ and $\beta_{\tau 2}^{*(j)}$ both between $\beta_{\tau 0}^{(j)}$ and $\beta_\tau^{(j)}$. Then, similar to (S.20) and (S.21), by Lemma A.1 and the ergodic theorem, together with $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = O_p(1)$ and $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$, it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \psi_\tau(\varepsilon_{t+k,\tau}) \zeta_{3nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) [1 - 2I(\varepsilon_t < b_\tau)] = o_p(1),$$

and

$$\begin{aligned}
& E \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \psi_\tau(\varepsilon_{t+k,\tau}) \zeta_{4nt}(\theta_\tau, \theta) [1 - 2I(\varepsilon_t < b_\tau)] \right| \\
& \qquad \leqslant \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} E \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{4nt}(\theta_\tau, \theta)| = O(n^{-1/2}),
\end{aligned}$$

which implies

$$\sum_{t=1}^{n-k} Z_{2nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) = o_p(1). \tag{S.61}$$

Similarly, using the Taylor expansion in (S.60), together with Lemma A.1 and Assumption 2, we can show that

$$\begin{aligned}
& E \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} \left| \sum_{t=1}^{n-k} Z_{3nt}(\theta_\tau, \theta) \right| \\
& \leqslant \frac{2}{\sqrt{n}} E \sum_{t=1}^{n-k} \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{1nt}(\theta_\tau, \theta)| I \left(|\varepsilon_t - b_\tau| \leqslant \sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{1nt}(\theta_\tau, \theta)| \right) \\
& \leqslant \frac{4 \sup_{x \in \mathbb{R}} f(x)}{\sqrt{n}} \sum_{t=1}^{n-k} E \left(\sup_{\theta_\tau \in \Theta_{\tau n}, \theta \in \Theta_n} |\zeta_{1nt}(\theta_\tau, \theta)| \right)^2 = O(n^{-1/2}),
\end{aligned}$$

and as a result,

$$\sum_{t=1}^{n-k} Z_{3nt}(\widehat{\theta}_{\tau n}, \widetilde{\theta}_n) = o_p(1). \tag{S.62}$$

Combining (S.58), (S.59), (S.61) and (S.62), we have

$$\sum_{t=k+1}^{n} \mathcal{E}_{2nt} = o_p(1). \tag{S.63}$$

Now we consider $\sum_{t=k+1}^{n} \mathcal{E}_{3nt}$. Similar to the proof of (S.44), for any $u, v \in \mathbb{R}^{p+q+1}$, define $\varphi_t(u,v) = \{I(\varepsilon_t < b_\tau) - I[\varepsilon_t < \widetilde{b}_t(u,v)]\} [|\widetilde{\varepsilon}_{t-k,\tau}(u,v)| - |\varepsilon_{t-k,\tau}|]$, where $\widetilde{\varepsilon}_{t,\tau}(u,v) = \left[y_t - (\theta_{\tau 0} + n^{-1/2}u)'\widetilde{z}_t(\theta_0 + n^{-1/2}v)\right] \widetilde{h}_t^{-1}(\theta_0 + n^{-1/2}v)$. Then, for any $M > 0$, we can readily verify that

$$\sup_{\|u\|,\|v\| \leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \{\varphi_t(u,v) - E[\varphi_t(u,v)|\mathcal{F}_{t-1}]\} \right| = o_p(1)$$

and

$$\sup_{\|u\|,\|v\| \leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} E[\varphi_t(u,v)|\mathcal{F}_{t-1}] \right| = o_p(1),$$

which yields

$$\sum_{t=k+1}^{n} \mathcal{E}_{3nt} = o_p(1). \tag{S.64}$$

Therefore, combining (S.43), (S.44), (S.63) and (S.64), we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \psi_\tau(\widehat{\varepsilon}_{t,\tau}) |\widehat{\varepsilon}_{t-k,\tau}| = & \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \psi_\tau(\varepsilon_{t,\tau}) |\varepsilon_{t-k,\tau}| \\ & - f(b_\tau) \left[d_{1k}' \sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) + b_\tau d_{2k}' \sqrt{n}(\widetilde{\theta}_n - \theta_0) \right] + o_p(1). \end{aligned} \tag{S.65}$$

Finally, by the law of large numbers and a proof similar to that for (S.58), we can show that

$$|\widehat{\mu}_{a,\tau} - \mu_{a,\tau}| = \left| \frac{1}{n} \sum_{t=1}^{n} (|\widehat{\varepsilon}_{t,\tau}| - |\varepsilon_{t,\tau}|) \right| + o_p(1) \leqslant \frac{1}{n} \sum_{t=1}^{n} |\widehat{\varepsilon}_{t,\tau} - \varepsilon_{t,\tau}| + o_p(1) = o_p(1),$$

and then,

$$\begin{aligned} \widehat{\sigma}_{a,\tau}^2 &= \frac{1}{n} \sum_{t=1}^{n} (|\widehat{\varepsilon}_{t,\tau}| - \widehat{\mu}_{a,\tau})^2 = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon}_{t,\tau}^2 - \mu_{a,\tau}^2 + o_p(1) \\ &= \frac{1}{n} \sum_{t=1}^{n} (\widehat{\varepsilon}_{t,\tau}^2 - \varepsilon_{t,\tau}^2) + \sigma_{a,\tau}^2 + o_p(1) \\ &= \sigma_{a,\tau}^2 + o_p(1), \end{aligned}$$

which, together with (S.65), (S.35) and (S.36), yields

$$r_{k,\tau} = \frac{1}{\sqrt{(\tau-\tau^2)\sigma_{a,\tau}^2}} \cdot \frac{1}{n}\sum_{t=k+1}^{n} \left\{ \psi_\tau(\varepsilon_{t,\tau}) \left(|\varepsilon_{t-k,\tau}| - d_{1k}'\Omega_2^{-1}\frac{z_t}{h_t} \right) \right.$$
$$\left. + b_\tau f(b_\tau) \left(d_{2k}' - d_{1k}'\Omega_2^{-1}\Gamma_2 \right) J^{-1} \frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} \right\} + o_p(n^{-1/2}). \tag{S.66}$$

Consequently, for $R = (r_{1,\tau}, \ldots, r_{K,\tau})'$, we have

$$R = \frac{1}{\sqrt{(\tau-\tau^2)\sigma_{a,\tau}^2}} \cdot \frac{1}{n}\sum_{t=k+1}^{n} \left\{ \psi_\tau(\varepsilon_{t,\tau}) \left(\epsilon_{t-1} - D_1\Omega_2^{-1}\frac{z_t}{h_t} \right) \right.$$
$$\left. + b_\tau f(b_\tau) \left(D_2 - D_1\Omega_2^{-1}\Gamma_2 \right) J^{-1} \frac{1-|\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} \right\} + o_p(n^{-1/2}), \tag{S.67}$$

where $\epsilon_{t-1} = (|\varepsilon_{t-1,\tau}|, \ldots, |\varepsilon_{t-K,\tau}|)'$ and $D_i = (d_{i1}, \ldots, d_{iK})'$ for $i = 1$ and 2. Applying the central limit theorem and the Cramér-Wold device, we have $\sqrt{n}R \to_d N(0, \Sigma_4)$.

To prove this theorem, it remains to show that Σ_4 is positive definite. Note that Σ_4 is the covariance matrix of $(\tau-\tau^2)^{-1/2}\sigma_{a,\tau}^{-1}(s_{1t}V_{1t} + s_{2t}V_{2t})$, where $s_{1t} = \psi_\tau(\varepsilon_{t,\tau})$, $s_{2t} = 1 - |\varepsilon_t|$,

$$V_{1t} = \epsilon_{t-1} - D_1\Omega_2^{-1}\frac{z_t}{h_t} \quad \text{and} \quad V_{2t} = b_\tau f(b_\tau)\left(D_2 - D_1\Omega_2^{-1}\Gamma_2 \right) J^{-1}\frac{1}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}.$$

Suppose that Σ_4 is singular. Then, there exists $\lambda \in \mathbb{R}^K$ such that $\lambda \neq 0$ and

$$s_{1t}\lambda' V_{1t} + s_{2t}\lambda' V_{2t} = 0 \quad \text{a.s.} \tag{S.68}$$

Since $s_{1t} = \psi_\tau(\varepsilon_{t,\tau}) = \tau - I(\varepsilon_t - b_\tau < 0) \neq 0$ for $\tau \in (0,1)$, (S.68) can be written as

$$\lambda' V_{1t} = -\frac{s_{2t}}{s_{1t}}\lambda' V_{2t} \quad \text{a.s.}$$

Note that s_{1t} and s_{2t} are independent of $\mathcal{F}_{t-1}$, and V_{1t} and V_{2t} are measurable with respect to $\mathcal{F}_{t-1}$. Taking the expectation conditional on $\mathcal{F}_{t-1}$ on both sides, we have

$$\lambda' V_{1t} = c\lambda' V_{2t} \quad \text{a.s.,}$$

where

$$c = -E\left(\frac{s_{2t}}{s_{1t}}\right) = E\left[\frac{|\varepsilon_t| - 1}{\tau - I(\varepsilon_t - b_\tau < 0)}\right].$$

As a result, (S.68) implies that

$$(cs_{1t} + s_{2t})\lambda' V_{2t} = 0 \quad \text{a.s.} \tag{S.69}$$

Define the $(p+q+1)\times 1$ constant vector

$$\mu = b_\tau f(b_\tau) J^{-1}\left(D_2 - D_1\Omega_2^{-1}\Gamma_2\right)'\lambda.$$

Then, $\lambda' V_{2t} = \mu' h_t^{-1}[\partial h_t(\theta_0)/\partial\theta]$, and (S.69) can be written as

$$(cs_{1t}+s_{2t})\frac{\mu'}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} = 0 \quad \text{a.s.} \tag{S.70}$$

Suppose that $\mu = 0$. Then, it follows from (S.68) that

$$\lambda' V_{1t} = 0 \quad \text{a.s.} \tag{S.71}$$

By a method similar to that for the proof of Theorem 8.2 in Francq and Zakoian (2010), we can show that (S.71) is impossible; we will prove this result at the end of the proof of this theorem.

Suppose that $\mu \neq 0$. From the proof of Theorem 2.2 in Francq and Zakoian (2004), by Assumption 1(iii), the matrix J must be positive definite: i.e., for $\mu \neq 0$, we have

$$P\left\{\frac{\mu'}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta} \neq 0\right\} > 0.$$

This, together with (S.70) and the independence of $cs_{1t}+s_{2t}$ and $\mu' h_t^{-1}[\partial h_t(\theta_0)/\partial\theta]$, implies that

$$cs_{1t}+s_{2t} = 0 \quad \text{a.s.}$$

That is, $|\varepsilon_t| - 1 = c\tau - cI(\varepsilon_t < b_\tau)$ almost surely, which is impossible due to the almost everywhere positiveness and differentiability of the density $f(\cdot)$ on a fixed small interval around b_τ. Therefore, Σ_4 is nonsingular.

Finally, we prove that (S.71) is impossible. Suppose that (S.71) holds. Denote $\lambda = (\lambda_1,\ldots,\lambda_K)'$ and define the $(p+q+1)\times 1$ constant vector

$$\gamma = (\gamma_1,\ldots,\gamma_{p+q+1})' = \Omega_2^{-1}D_1'\lambda.$$

Then, (S.71) can be written as

$$\lambda'\epsilon_{t-1} - \gamma'\frac{z_t}{h_t} = 0 \quad \text{a.s.} \tag{S.72}$$

Note that $\gamma \neq 0$. Otherwise, $\lambda'\epsilon_{t-1}$=0 almost surely, which implies that there exists $j \in \{1,\ldots,K\}$ such that $\lambda_j \neq 0$ and $|\varepsilon_{t-j,\tau}| = -\lambda_j^{-1}\sum_{i=1,i\neq j}^K \lambda_i|\varepsilon_{t-i,\tau}|$. By the independence of $|\varepsilon_{t-1,\tau}|,\ldots,|\varepsilon_{t-K,\tau}|$, we then have that $|\varepsilon_{t,\tau}|$ is degenerate, which is true if and only if ε_t is degenerate. This contradicts Assumption 1(ii). Thus, $\gamma \neq 0$.

By (S.72) and the positiveness of h_t, we have

$$h_t\lambda'\epsilon_{t-1} - \gamma' z_t = 0 \quad \text{a.s.} \tag{S.73}$$

For notational simplicity, we denote by $R_1, R_2, \dots$ random variables measurable with respect to $\mathcal{F}_{t-2}$. Then we have $h_t = \alpha_{01}h_{t-1}|\varepsilon_{t-1}| + R_1$, $\lambda'\epsilon_{t-1} = \lambda_1|\varepsilon_{t-1,\tau}| + R_2$, and $\gamma' z_t = \gamma_2|y_{t-1}| + R_3$. As a result, it follows from (S.73) that

$$\lambda_1\alpha_{01}h_{t-1}|\varepsilon_{t-1}||\varepsilon_{t-1,\tau}| + (\alpha_{01}R_2 - \gamma_2)h_{t-1}|\varepsilon_{t-1}| + \lambda_1 R_1|\varepsilon_{t-1,\tau}| + R_4 = 0 \quad \text{a.s.} \tag{S.74}$$

If $\lambda_1\alpha_{01} \neq 0$, then (S.74) implies that $(|\varepsilon_{t-1}| - R_5)(|\varepsilon_{t-1,\tau}| - R_6) = 0$ almost surely, which is impossible since ε_t is non-degenerate. Thus, $\lambda_1\alpha_{01} = 0$ must hold.

If $\lambda_1 = 0$, then it follows from (S.74) that $(\alpha_{01}R_2 - \gamma_2)h_{t-1}|\varepsilon_{t-1}| + R_4 = 0$ almost surely. Taking the expectation conditional on $\mathcal{F}_{t-2}$, we have $(\alpha_{01}R_2 - \gamma_2)h_{t-1} + R_4 = 0$ almost surely. In view of the positiveness of h_{t-1}, it follows that

$$(\alpha_{01}R_2 - \gamma_2)(|\varepsilon_{t-1}| - 1) = 0 \quad \text{a.s.}$$

Since ε_t is non-degenerate and $|\varepsilon_{t-1}| - 1$ is independent of $\alpha_{01}R_2 - \gamma_2$, this implies that $\alpha_{01}R_2 - \gamma_2 = 0$ almost surely. Note that $R_2 = \sum_{i=2}^{K}\lambda_i|\varepsilon_{t-i,\tau}|$, where at least one of $\lambda_2, \dots, \lambda_K$ is nonzero. By an argument used earlier, we have $P(R_2 \neq 0) > 0$. Consequently, $\alpha_{01} = \gamma_2 = 0$. However, from the second paragraph in Section 2.1, we assume $\alpha_{01} \geqslant \underline{w} > 0$. The conclusion follows. □

Proof of Theorem 4. Similar to (S.43), we have

$$\begin{aligned}\frac{1}{\sqrt{n}}&\sum_{t=k+1}^{n}\omega_t\psi_\tau(\widehat{\varepsilon}^*_{t,\tau})|\widehat{\varepsilon}^*_{t-k,\tau}| \\ &= \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\omega_t\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}| + \sum_{t=k+1}^{n}\mathcal{E}^*_{1nt} + \sum_{t=k+1}^{n}\mathcal{E}^*_{2nt} + \sum_{t=k+1}^{n}\mathcal{E}^*_{3nt},\end{aligned} \tag{S.75}$$

where

$\mathcal{E}^*_{1nt} = n^{-1/2}\omega_t[\psi_\tau(\widehat{\varepsilon}^*_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})]|\varepsilon_{t-k,\tau}|$, $\mathcal{E}^*_{2nt} = n^{-1/2}\omega_t\psi_\tau(\varepsilon_{t,\tau})(|\widehat{\varepsilon}^*_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|)$, and $\mathcal{E}^*_{3nt} = n^{-1/2}\omega_t[\psi_\tau(\widehat{\varepsilon}^*_{t,\tau}) - \psi_\tau(\varepsilon_{t,\tau})](|\widehat{\varepsilon}^*_{t-k,\tau}| - |\varepsilon_{t-k,\tau}|)$.

Note that, from (S.39) and (S.42), $\sqrt{n}(\widetilde{\theta}^*_n - \theta_0) = O^*_p(1)$ and $\sqrt{n}(\widehat{\theta}^*_{\tau n} - \theta_{\tau 0}) = O^*_p(1)$. As a result, by methods similar to (S.44), (S.63) and (S.64), respectively, we can show that

$$\sum_{t=k+1}^{n}\mathcal{E}^*_{1nt} = -f(b_\tau)\left[d'_{1k}\sqrt{n}(\widehat{\theta}^*_{\tau n} - \theta_{\tau 0}) + b_\tau d'_{2k}\sqrt{n}(\widetilde{\theta}^*_n - \theta_0)\right] + o^*_p(1),$$

and

$$\sum_{t=k+1}^{n} \mathcal{E}_{int}^{*} = o_p^{*}(1), \quad i = 2 \text{ and } 3,$$

where $d_{1k} = E(h_t^{-1}|\varepsilon_{t-k,\tau}|z_t)$ and $d_{2k} = E(h_t^{-1}|\varepsilon_{t-k,\tau}| \sum_{j=1}^{p} \beta_{0j}\partial h_{t-j}(\theta_0)/\partial\theta)$ are defined as in (S.44). This, in conjunction with (S.75) and (S.65), yields the Bahadur representation of

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \omega_t \psi_\tau(\widehat{\varepsilon}_{t,\tau}^{*})|\widehat{\varepsilon}_{t-k,\tau}^{*}| - \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \psi_\tau(\widehat{\varepsilon}_{t,\tau})|\widehat{\varepsilon}_{t-k,\tau}| \\
= \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} (\omega_t - 1)\psi_\tau(\varepsilon_{t,\tau})|\varepsilon_{t-k,\tau}| \\
- f(b_\tau)\left[d_{1k}'\sqrt{n}(\widehat{\theta}_{\tau n}^{*} - \widehat{\theta}_{\tau n}) + b_\tau d_{2k}'\sqrt{n}(\widetilde{\theta}_{n}^{*} - \widetilde{\theta}_{n})\right] + o_p^{*}(1),
\end{aligned}$$

and hence

$$\begin{aligned}
R^{*} - R = \frac{1}{\sqrt{(\tau - \tau^2)\sigma_{a,\tau}^2}} \cdot \frac{1}{n} \sum_{t=k+1}^{n} (\omega_t - 1)\Bigg\{\psi_\tau(\varepsilon_{t,\tau})\left(\epsilon_{t-1} - D_1\Omega_2^{-1}\frac{z_t}{h_t}\right) \\
+ b_\tau f(b_\tau)\left(D_2 - D_1\Omega_2^{-1}\Gamma_2\right) J^{-1}\frac{1 - |\varepsilon_t|}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}\Bigg\} + o_p^{*}(n^{-1/2}),
\end{aligned}$$

where $\epsilon_{t-1} = (|\varepsilon_{t-1,\tau}|, \ldots, |\varepsilon_{t-K,\tau}|)'$ and $D_i = (d_{i1}, \ldots, d_{iK})'$ for $i = 1$ and 2. Thus, we complete the proof by applying Lindeberg's central limit theorem and the Cramér-Wold device. □

Proof of Corollary 1. The proof follows the same lines as that of Theorem 1, while the corresponding $L_{1n}(u)$ and $L_{2n}(u)$ are defined with $\widetilde{h}_t^{-1}$ replaced by one; consequently, all the $A_{int}(\theta)$'s and $B_{int}(\theta)$'s are defined with all $\widetilde{h}_t^{-1}(\theta)$, $h_t^{-1}(\theta)$ and h_t^{-1} replaced by one. Note that without these denominators, Lemma A.1 cannot be applied as in the proof of Theorem 1 in some intermediate steps, and additional moment conditions on x_t will be needed. The highest moment condition, $E|x_t|^{4+\iota_0}$ for some $\iota_0 > 0$, is required for the proof of the counterpart of (S.31), where, correspondingly, $\eta_t(v) = \int_0^{\xi_{1nt}(\theta_0 + n^{-1/2}v)} I_t^{*}(s)ds$, with ξ_{1nt} and $I_t^{*}(s)$ defined as in the proof of Theorem 1. The corresponding proof is straightforward by the Hölder inequality. □

Proof of Corollary 2 and Equation (2.6). Since $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = O_p(1)$ and $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0}) = O_p(1)$, Corollary 2 follows directly from Lemma S.1 and the Taylor expansion.

Moreover, it can be readily shown that the sequence $\{X_n\}$ with $X_n = u_{n+1}'\sqrt{n}(\widetilde{\theta}_n - \theta_0) + z_{n+1}'\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_{\tau 0})$ is uniformly tight, which, combined with Corollary 2, implies that $o_p(|\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n)|) = o_p(n^{-1/2})$. Note that $b_\tau \neq 0$ if and only if $Q_\tau(y_{n+1}|\mathcal{F}_n) = \theta'_{\tau 0} z_{n+1} = b_\tau h_{n+1} \neq 0$, since $h_{n+1} \geqslant \underline{w} > 0$. If $b_\tau \neq 0$, then $T^{-1}(\cdot)$ is differentiable at $Q_\tau(y_{n+1}|\mathcal{F}_n)$, and hence

$$\begin{aligned} & T^{-1}[\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n)] - T^{-1}[Q_\tau(y_{n+1}|\mathcal{F}_n)] \\ &= \left.\frac{dT^{-1}(x)}{dx}\right|_{x=Q_\tau(y_{n+1}|\mathcal{F}_n)} \left[\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n)\right] + o_p(n^{-1/2}) \\ &= \frac{1}{2\sqrt{|b_\tau h_{n+1}|}} \left[u'_{n+1}(\widetilde{\theta}_n - \theta_0) + z'_{n+1}(\widehat{\theta}_{\tau n} - \theta_{\tau 0})\right] + o_p(n^{-1/2}). \end{aligned}$$

Since $\widehat{Q}_\tau(x_{n+1}|\mathcal{F}_n) = T^{-1}[\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n)]$ and $Q_\tau(x_{n+1}|\mathcal{F}_n) = T^{-1}[Q_\tau(y_{n+1}|\mathcal{F}_n)]$, we complete the proof of (2.6). □

Proof of Corollary 3. By methods similar to the proofs of Theorem 2 and Corollary 2, this corollary follows. □

References

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 32:633–655.

Berkes, I., Horváth, L., and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli*, 9:201–227.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78:1093–1125.

Davis, R. A., Knight, K., and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications*, 40:145–180.

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. (2010). GARCH *Models: Structure, Statistical Inference and Financial Applications.* John Wiley & Sons, Chichester, UK.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics*, 26:755–770.

Li, G., Leng, C., and Tsai, C.-L. (2014). A hybrid bootstrap approach to unit root tests. *Journal of Time Series Analysis*, 35:299–321.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21:255–285.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Zheng, Y., Li, Y., and Li, G. (2016). On Fréchet autoregressive conditional duration models. *Journal of Statistical Planning and Inference*, 175:51–66.