Supplementary material for "On The Least Squares Estimation of Multiple-Threshold-Variable Autoregressive Models"

The supplementary contains five parts. Section S.1 discusses information criteria for order selection and regime specification, and provides an example of regime coalescence. Section S.2 gives the extended weighted Nadaraya-Watson (WNW) and modified resampling algorithms with all the parameters estimated. In order to compare with the diminishing-threshold-effect framework in Chen et al. (2012), Section S.3 discusses on confidence interval construction for thresholds, provides theoretical support and gives simulation results for different distributions. Section S.4 and S.5 contain proofs of the main results and some auxiliary lemmas.

S.1 Information Criteria

In this section, we first discuss information criteria for order selection of threshold models, and the reason we adopt the AICu. Then, we use the AICu to specify the regime structure and give an example based on the data in Section 6.2.

S.1.1 Information Criteria for Threshold Models

For the 2-TAR model, in the estimation procedure, we use information criteria to determine the regime splitting and the order simultaneously. Specifically, the estimation of threshold and order selection interact in the following way:

1. For each threshold (r, s), select order using information criteria within some maximum order $p_{\max}$.
2. Repeat the above procedure in the grid search for each (r, s).
3. Obtain the smallest information criteria among grid search, then output the according (r, s) and AR coefficients with order p_j, $j = 1, \dots, 4$.

For the self-exciting threshold autoregressive (SETAR) model with one threshold variable, there is a moderate amount of literature proposing different information criteria and assessing their performance. See, e.g. Wong and Li (1998), De Gooijer (2001), Peña and Rodriguez (2005), among others. Commonly used ones include the AIC, BIC, AICc, AICu, among others. For the 2-TAR model, they write as follows.

$$\mathrm{AIC}(\{p_j\}) = \sum_{j=1}^{4} \left[n_j \log(\widehat{\sigma}_{jn}^2) + 2(p_j+1) \right],$$

$$\mathrm{BIC}(\{p_j\}) = \sum_{j=1}^{4} \left[n_j \log(\widehat{\sigma}_{jn}^2) + \log n_j (p_j+1) \right],$$

$$\mathrm{AICc}(\{p_j\}) = \sum_{j=1}^{4} \left[n_j \log(\widehat{\sigma}_{jn}^2) + \frac{2n_j(p_j+2)}{n_j-p_j-3} \right],$$

$$\mathrm{AICu}(\{p_j\}) = \sum_{j=1}^{4} \left[n_j \log(\widehat{\sigma}_{jn}^2) + \frac{2n_j(p_j+2)}{n_j-p_j-3} + n_j \log\left(\frac{n_j}{n_j-p_j-2} \right) \right],$$

where n_j, p_j and $\widehat{\sigma}_{jn}^2$ are the number of observations, order and estimated residual variance in the jth regime, respectively. From the literature, a good choice for order selection is the AICu, which has a competitive small sample performance and computational convenience. Let us now provide an additional argument in favour of the AICu for our case.

Now, simultaneous regime splitting and order selection tends to cause the problem of "small and overfitted regimes" because excessive regime splitting tends to be favoured by the information criteria. Specifically, consider the case in which one of the regimes, say the jth regime, is small in sample size and over-fitted. Then, the variance $\widehat{\sigma}_{jn}^2$ will be close to zero and approaching negative infinity after taking logarithm, implying that the first term in the information criteria, $\sum_{j=1}^4 n_j \log(\widehat{\sigma}_{jn}^2)$, will be large negatively, making this choice very competitive in the above estimation procedure. This problem is particular to threshold modeling, since the procedure involves simultaneous regime splitting and order selection.

One way to handle this problem is to penalize the possibility of getting a "small and overfitted" regime. Apparently the AIC and BIC cannot do that, while the AICc and AICu are workable, since the terms $\frac{n_j}{n_j-p_j-3}$ and $\frac{n_j}{n_j-p_j-2}$ are close to 1 for regimes with moderately large sample size but will be large for "small and overfitted" regimes. Between the AICc and AICu, the latter has heavier penalties.

Let's take a specific example to illustrate the potential disadvantage of the AIC. Consider the empirical application in Section 6.1. If we adopt the AIC in the estimation procedure, we finally obtain the estimate $\hat{r}_n = -0.212, \hat{s}_n = 0.571$ with the smallest the AIC, which is -222.772. In the four regimes, we have

$$\begin{aligned}
&p_1 = 2, \quad p_2 = 10, \quad p_3 = 10, \quad p_4 = 9;\\
&n_1 = 77, \quad n_2 = 98, \quad n_3 = 12, \quad n_4 = 90;\\
&\widehat{\sigma}_1 = 0.678, \quad \widehat{\sigma}_2 = 0.717, \quad \widehat{\sigma}_3 = 0.003, \quad \widehat{\sigma}_4 = 0.847.
\end{aligned}$$

The AICs in the four regimes are -53.858, -43.201, -115.767 and -9.948, respectively. Clearly Regime 3 is a small and overfitted regime, and its small variance is largely negative after taking the logarithm. We can see that Regime 3 contributes a lot to the total AIC, although it has the fewest observations, making this choice very competitive in the estimation procedure, which turns out to be our final estimate. To compare, the estimation using the AICu in Section 6.1 is less prone to this overfitting phenomenon.

Thus, as an interim choice, we decide to use the AICu in the estimation procedure. However, further research is needed, especially when the threshold variables are multiple or exogenous.

S.1.2 An Example of Regime Coalescence by the AICu

In this part, we consider the regime structure specification using information criteria. For the self-exciting threshold autoregressive (SETAR) model with one threshold variable, there is a moderate number of alternatives, mostly to do with information criteria of some kind, to select the number of regimes, e.g. Gonzalo and Pitarakis (2002), Hamaker (2009), among others. However, as far as we know, it is still a blank for the multiple-threshold-variable model as well as for the exogenous-threshold-variable model. As an interim solution, we use the AICu for regime selection.

In the empirical application in Section 6.2, we note that the coefficients of Regimes 2 and 3 are close, suggesting potential regime coalescence. Thus, we reanalyze this data. Recalling that a time series $\{y_t\}$ is said to be a pth order 2-TAR model if it satisfies

$$y_t = \sum_{j=1}^{4} (\boldsymbol{\beta}_j' \mathbf{y}_{t-1} + \sigma_j \epsilon_t) I_{jt}(r, s), \tag{S.1.1}$$

where $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})' \in \mathbb{R}^{p+1}$, $\sigma_j > 0$. There are potentially 8 additional regime coalescences.

$$\begin{aligned}
S1: &\quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2, \quad \sigma_1 = \sigma_2. \\
S2: &\quad \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3, \quad \sigma_2 = \sigma_3. \\
S3: &\quad \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4, \quad \sigma_3 = \sigma_4. \\
S4: &\quad \boldsymbol{\beta}_4 = \boldsymbol{\beta}_1, \quad \sigma_4 = \sigma_1. \\
S5: &\quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3, \quad \sigma_1 = \sigma_2 = \sigma_3. \\
S6: &\quad \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4, \quad \sigma_2 = \sigma_3 = \sigma_4. \\
S7: &\quad \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5, \quad \sigma_3 = \sigma_4 = \sigma_5. \\
S8: &\quad \boldsymbol{\beta}_4 = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2, \quad \sigma_4 = \sigma_1 = \sigma_2.
\end{aligned}$$

$S1$-$S4$ are structures with 3 regimes and $S5$-$S8$ are structures with 2 regimes. And let's call the 4-regime structure in (S.1.1) $S0$. Then, our modelling procedures is modified into:

1. For $j = 0, \ldots, 8$, under the regime structure Sj, implement the estimation procedure in Section S.1.1.

2. Choose the smallest the AICu among the nine Sj, and output the according Sj and estimated parameters.

Following this procedure, we reanalyze the SHI stock data, and the estimation results could be found in Section 6.2.

S.2 Extended Algorithms

In Section 4, we give **Algorithm B** and **C** to estimate $F_1(\gamma|r_0)$ conditional on the assumption that all other parameters are known. In practice, we need to first estimate these parameters, including $\boldsymbol{\theta}_0$, σ_{j0}, $\pi_1(r_0)$, $f_\epsilon(\cdot)$ and $F_\epsilon(\cdot)$. Given the sample $\mathcal{X}$, following the estimation procedure in Li and Ling (2012), we obtain their consistent estimator, denoting as $\widehat{\boldsymbol{\theta}}_n$, $\widehat{\sigma}_{jn}$, $\widehat{\pi}_1(\hat{r}_n)$, $\widehat{f}_\epsilon(\cdot)$ and $\widehat{F}_\epsilon(\cdot)$. Meanwhile, we obtain the residuals $\{\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n\}$. Thus, we have

$$\widehat{\gamma}_t^{(1)} = \widehat{\xi}_t^{(1,2)} I(w_{t-1} > \hat{s}_n) + \widehat{\xi}_t^{(4,3)} I(w_{t-1} \leq \hat{s}_n), \tag{S.2.1}$$

with

$$\widehat{\xi}_t^{(i,j)} = [(\widehat{\boldsymbol{\beta}}_{in} - \widehat{\boldsymbol{\beta}}_{jn})'\mathbf{y}_{t-1}]^2 + 2\widehat{\sigma}_{in}\hat{\epsilon}_t(\widehat{\boldsymbol{\beta}}_{in} - \widehat{\boldsymbol{\beta}}_{jn})'\mathbf{y}_{t-1}, \quad i, j = 1, ..., 4.$$

Then, we give **Algorithm B′** and **C′**, extending **Algorithm B** and **C** by substituting all parameters by their consistent estimates.

Algorithm B′:

Step B′.1 For a fixed γ,

$$\widetilde{F}_1^{\text{WNW}}(\gamma|\hat{r}_n, \mathcal{X}) = \frac{\sum_{t=1}^n p_t(\hat{r}_n) K_h(\hat{r}_n - z_{t-1}) I(\widehat{\gamma}_t^{(1)} \leq \gamma)}{\sum_{t=1}^n p_t(\hat{r}_n) K_h(\hat{r}_n - z_{t-1})}. \tag{S.2.2}$$

Step B′.2 $\widetilde{F}_1^{\text{WNW}}(\cdot|r_0, \mathcal{X})$ is defined as the cumulative distribution function of a discrete distribution taking values at $\{\widehat{\gamma}_t^{(1)} : t = 1, \ldots, n\}$ with the according cumulative probability $\widetilde{F}_1^{\text{WNW}}(\widehat{\gamma}_t^{(1)}|\hat{r}_n, \mathcal{X})$. Draw a random sample from this discrete distribution, and denote it as $\zeta_1^{(1),\text{WNW}}$.

Algorithm C′:

Step C′.1 For each $i = 1, ..., n$, set $\boldsymbol{q}_i = (y_i, ..., y_{i-p+1})'$ and generate a sample $\{y_1^{(i)}, \ldots, y_{2-d_1-p}^{(i)}\}$ in the following way: first let $(y_{1-d_1}^{(i)}, ..., y_{2-d_l-p}^{(i)})' = \boldsymbol{q}_i$ and $y_{2-d_1}^{(i)} = l_{\boldsymbol{q}_i}^{-1}(\hat{r}_n)$; draw $\{\hat{\epsilon}_{3-d_1}^{(i)}, ..., \hat{\epsilon}_1^{(i)}\}$ independently from $\widehat{F}_\epsilon(\cdot)$ and generate $y_{3-d_1}^{(i)}, ..., y_1^{(i)}$ by iterating model (2.1) based on the initial values $\{y_{2-d_1}^{(i)}, y_{1-d_1}^{(i)}, ..., y_{2-d_1-p}^{(i)}\}$ and $\boldsymbol{\theta}_0$ and σ_{j0} being replaced by $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\sigma}_{jn}$. Then, calculate $\widehat{\gamma}_2^{(1)}$ in (S.2.1) based on the sample $\{y_1^{(i)}, \ldots, y_{2-d_1-p}^{(i)}\}$ and denote it as ζ_i.

Step C′.2 Replace the $\boldsymbol{\theta}_0, \boldsymbol{\sigma}_0, f_\epsilon(\cdot)$ by their estimate $\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\sigma}}_n, \widehat{f}_\epsilon(\cdot)$ in (4.2), and calculate $\widehat{\pi}_1(\hat{r}_n|\boldsymbol{q}_i)$'s for $i = 1, ..., n$. Draw a U from the discrete distribution: $\mathbb{P}(U = i|\mathcal{X}) = \widehat{\pi}_1(\hat{r}_n|\boldsymbol{q}_i)/\{\sum_{j=1}^n \widehat{\pi}_1(\hat{r}_n|\boldsymbol{q}_j)\}$, conditionally independent of all $\{\hat{\epsilon}_{3-d_1}^{(i)}, ..., \hat{\epsilon}_1^{(i)}\}$ given $\mathcal{X}$.

Step C′.3 Obtain $\zeta_1^{(1),\text{RS}} = \zeta_U$.

S.3 Comparison of confidence intervals for thresholds

One important topic is to compare the fixed-threshold-effect framework and the diminishing-threshold-effect framework. Li et al. (2019) and Yu and Phillips (2018) have given some results on the connection of the asymptotics for these two frameworks under the single threshold variable setting, which could be generalized to the multiple threshold variable setting. In this section, we focus on confidence interval (CI) construction for thresholds in our paper and Chen et al. (2012). However, since Chen et al. (2012) did not detail how to construct CIs, we will first introduce the method and its theoretical support following the logic and routine in Hansen (2000). Then, some simulation results will be given to compare the empirical performance of the CI construction methods. Since Chen et al. (2012) focused on models with two threshold variables and homoskedastic errors, we pay attention to this specific case for comparison.

S.3.1 CI construction under the diminishing-threshold-effect framework

Under the single threshold variable setting, Hansen (2000) constructed CI based on the likelihood ratio test statistic. He did not use the commonly used Wald statistic, since the fact that the asymptotic distributions depend on unknown parameters may lead to poor finite sample performance. Generalizing his method, CIs can be constructed for models with multiple threshold variables. However, Chen et al. (2012) did not demonstrate CI construction issues. Thus, we first give the details of CI construction for the models considered in Chen et al. (2012), i.e., models with two threshold variables and homoskedastic errors.

As shown in Section 2, based on the sum-of-squared-error function $L_n(\boldsymbol{\theta})$, we could get the concentrated one $L_n^*(\boldsymbol{\tau})$. Define the likelihood ratio test statistic as

$$LR_n(\boldsymbol{\tau}) = n\frac{L_n^*(\boldsymbol{\tau}) - L_n^*(\hat{\tau})}{L_n^*(\hat{\tau})}.$$

Then, the asymptotics of $LR_n(\boldsymbol{\tau})$ is given as below.

Lemma S.3.1. *Under conditions of Theorem 1 in Chen et al. (2012), we have*

$$LR_n(\boldsymbol{\tau}_0) \to_d \phi = \max_{r,s\in\mathbb{R}} \left[2W_1(r) - |r| + 2W_2(s) - |s|\right],$$

and $W_1(r)$ and $W_2(s)$ are each a two-sided Brownian motion defined as

$$W_1(r) = \begin{cases} \Lambda_{11}(-r) & \text{if } r < 0 \\ 0 & \text{if } r = 0 \\ \Lambda_{12}(r) & \text{if } r > 0 \end{cases}, \quad W_2(s) = \begin{cases} \Lambda_{21}(-s) & \text{if } s < 0 \\ 0 & \text{if } s = 0 \\ \Lambda_{22}(s) & \text{if } s > 0 \end{cases},$$

where $\Lambda_{11}(r)$, $\Lambda_{12}(r)$, $\Lambda_{21}(s)$ and $\Lambda_{22}(s)$ are four independent standard Brownian motions on $[0, \infty)$. And ϕ has the distribution function

$$\mathbb{P}(\phi < x) = \int_0^x e^{-z/2}[z(1 + e^{-z/2}) + 4(e^{-z/2} - 1)]dz.$$

Proof. The limiting distribution of ϕ follows from Theorem 1 in Chen et al. (2012). As for the distribution function, first define $\phi_1 = \max_{r\in\mathbb{R}}[2W_1(r) - |r|]$ and $\phi_2 = \max_{s\in\mathbb{R}}[2W_2(s) - |s|]$. Thus, ϕ_1 and ϕ_2 are i.i.d. By Theorem 2 in Hansen (2000), we have that $\mathbb{P}(\phi_1 \le x) = (1 - e^{-x/2})^2$ for $x \ge 0$. It follows that its probability density function is $f_{\phi_1}(x) = (1 - e^{-x/2})e^{-x/2}$. Thus, by a convolution argument, we have

$$
\begin{aligned}
f_\phi(z) &= \int_0^z f_{\phi_1}(x) f_{\phi_1}(z-x)dx \\
&= \int_0^z (1 - e^{-x/2})e^{-x/2}(1 - e^{-(z-x)/2})e^{-(z-x)/2}dx \\
&= e^{-z/2}\left[(1 + e^{-z/2})x + 2e^{-x/2} - 2e^{-(z-x)/2}\right]|_0^z \\
&= e^{-z/2}[z(1 + e^{-z/2}) + 4(e^{-z/2} - 1)],
\end{aligned}
$$

and the distribution function follows. □

Since the distribution is available in its closed form, critical values can be easily calculated and p-value could be obtained. Selected critical values $c_\phi(C)$ at the significance level C (e.g. C=0.95) are reported in Table 1.

Table 1: Asymptotic critical values of ϕ.

C	0.80	0.85	0.90	0.95	0.99
$c_\phi(C)$	8.33	9.13	10.21	11.98	15.86

Then, under the diminishing-threshold-effect framework and with the conditions in Lemma S.3.1, the set

$$\{\boldsymbol{\tau} : LR_n(\boldsymbol{\tau}) \le C\} \tag{S.3.1}$$

is the asymptotic C-level confidence regions based on the likelihood ratio test statistic $LR_n(\boldsymbol{\tau})$. A graphical method to find the region is to draw $LR_n(\boldsymbol{\tau})$ against $\boldsymbol{\tau}$ and draw a flat plane at $c_\phi(C)$. Thus, the confidence regions are joint instead of individual for thresholds. Besides, the confidence region may be disconnected. To solve this for the single threshold case, Hansen (1997) proposed a more conservative procedure by defining the smallest convexified interval that incorporates all the confidence regions. This procedure can be generalized to the 2-TAR model here by changing the convexified interval to a convexified rectangle.

When the threshold effect is relatively large or fixed, the situation becomes more complicated. Theorem 3 in Hansen (2000) considers the fixed-threshold-effect framework. The theorem shows that in the special case of iid Gaussian errors, the likelihood ratio test is asymptotically conservative. For generic cases, the paper claimed that "Unfortunately, we do not know if Theorem 3 generalizes to the case of non-normal errors or regressors that are not strictly exogenous. The proof of Theorem 3 relies on the Gaussian error structure and it is not clear how the theorem would generalize." As far as we know, there are still no breakthroughs on this problem.

S.3.2 Simulation results on CI under two frameworks

In this section, we give simulation results on CI construction for 2-TAR model with homoskedastic errors, since it is the focus of Chen et al. (2012). And we will evaluate the empirical performance of CIs construction under the two frameworks. We consider different distributions for errors, including normal, t, uniform, skew normal and skew t distribution. For the two skew distributions, we follow the notation in Azzalini and Capitanio (2014). We consider two 2-TAR models with relatively large and relatively small threshold effects respectively.

Specifically, the two models are

$$y_t = \begin{cases} -0.6y_{t-1} + 0.7y_{t-2} + \epsilon_t, & \text{if } y_{t-2} > 0,\ y_{t-1} > 0, \\ 0.7y_{t-1} - 0.6y_{t-2} + \epsilon_t, & \text{if } y_{t-2} \le 0,\ y_{t-1} > 0, \\ -0.7y_{t-1} + 0.6y_{t-2} + \epsilon_t, & \text{if } y_{t-2} \le 0,\ y_{t-1} \le 0, \\ 0.6y_{t-1} - 0.7y_{t-2} + \epsilon_t, & \text{if } y_{t-2} > 0,\ y_{t-1} \le 0, \end{cases} \tag{S.3.2}$$

$$y_t = \begin{cases} -0.2y_{t-1} + 0.4y_{t-2} + \epsilon_t, & \text{if } y_{t-2} > 0,\ y_{t-1} > 0, \\ 0.4y_{t-1} - 0.2y_{t-2} + \epsilon_t, & \text{if } y_{t-2} \le 0,\ y_{t-1} > 0, \\ -0.4y_{t-1} + 0.2y_{t-2} + \epsilon_t, & \text{if } y_{t-2} \le 0,\ y_{t-1} \le 0, \\ 0.2y_{t-1} - 0.4y_{t-2} + \epsilon_t, & \text{if } y_{t-2} > 0,\ y_{t-1} \le 0, \end{cases} \tag{S.3.3}$$

where in these two models, ϵ_t is i.i.d from $\mathcal{N}(0,1)$, $t(3)/\sqrt{3}$, $U(-\sqrt{3},\sqrt{3})$, $\mathcal{SN}(-1.26, 1.61, 5)$ and $\mathcal{ST}(0.82, 0.91, -2, 4)$, respectively. The parameters are set such that ϵ_t has 0 mean and unit variance. Let the sample sizes be $n = 600$, 900, and 1200, each with 1000 replications.

To evaluate the performance of CI construction, we assess the coverage probabilities, mean and standard deviation of the CI lengths under the two frameworks. Under the diminishing threshold effect framework, the coverage probabilities are considered for CIs by both the likelihood ratio test statistic and the conservative convexified region approach, which are indicated by I and II, respectively. As for CI lengths, since they are unavailable for method I, we use the lengths of two sides of the convexified rectangles in method II. Tables 2-11 report the results for model (S.3.2), while tables 12-21 report the results for model (S.3.3).

As we can see, for model (S.3.2) where the threshold effect is relatively large, the CIs under the fixed framework have quite accurate empirical significance levels. While the CIs under the diminishing framework generally miss the nominal levels and tend to be rather conservative in many cases. This conservative phenomenon tends to be more obvious for light tail distribution (uniform distribution). CIs under the fixed framework tend to have larger lengths and much smaller standard deviation, except for t distribution.

For model (S.3.3) where the threshold effect is relatively small, the empirical significance levels for CIs under the fixed framework tend to approach the nominal levels from below as n increases. Conservativeness under the diminishing framework persists, albeit less severely than in the large threshold effect model. Under both frameworks, the mean and standard deviations of CI lengths are much larger than those for model (S.3.2).

To summarize, under the fixed framework, CIs enjoy accurate coverage and stable lengths when the threshold effect is relatively large, while they tend to be non-conservative and might need a large sample size for satisfactory coverage when the threshold effect is relatively small. Under the diminishing framework, CIs tend to be conservative generally speaking, with a severity that tends to be lessened when the threshold effect is small. These findings are consistent with Hansen (2000). The choice between the two frameworks is an interesting and challenging problem and is beyond the scope of this paper. Interested readers can find discussions in Yu and Phillips (2018) and Li et al. (2019).

Table 2: Coverage probabilities for normal distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.906	0.897	0.914	0.886	0.910	0.900	0.982	0.988	0.988	0.993	0.995	0.995
0.950	0.955	0.953	0.960	0.934	0.958	0.953	0.996	0.996	0.994	1.000	0.998	0.997
0.990	0.993	0.990	0.994	0.986	0.990	0.995	0.999	1.000	0.999	1.000	1.000	1.000

Table 3: Mean and standard deviation of CI lengths for normal distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.078	0.051	0.038	0.068	0.045	0.033	0.071	0.045	0.033	0.057	0.037	0.027
	0.013	0.007	0.005	0.011	0.006	0.004	0.044	0.030	0.021	0.038	0.026	0.019
0.950	0.106	0.070	0.052	0.092	0.061	0.045	0.079	0.051	0.037	0.063	0.041	0.030
	0.018	0.010	0.006	0.015	0.008	0.005	0.048	0.032	0.023	0.040	0.028	0.019
0.990	0.177	0.116	0.087	0.152	0.101	0.075	0.098	0.062	0.045	0.076	0.050	0.036
	0.033	0.018	0.011	0.026	0.014	0.009	0.055	0.036	0.026	0.045	0.032	0.022

Table 4: Coverage probabilities for t distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.885	0.902	0.914	0.896	0.906	0.897	0.940	0.951	0.952	0.944	0.966	0.961
0.950	0.938	0.948	0.952	0.937	0.950	0.941	0.949	0.961	0.967	0.956	0.973	0.973
0.990	0.984	0.988	0.982	0.978	0.985	0.978	0.964	0.979	0.982	0.970	0.983	0.984

Table 5: Mean and standard deviation of CI lengths for t distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.083	0.052	0.039	0.068	0.045	0.032	0.071	0.048	0.034	0.058	0.040	0.029
	0.072	0.020	0.028	0.031	0.029	0.009	0.071	0.032	0.023	0.052	0.028	0.021
0.950	0.115	0.074	0.055	0.094	0.063	0.045	0.081	0.054	0.038	0.065	0.045	0.033
	0.089	0.029	0.034	0.045	0.042	0.014	0.077	0.036	0.025	0.057	0.030	0.022
0.990	0.197	0.129	0.096	0.161	0.109	0.079	0.104	0.067	0.047	0.079	0.054	0.039
	0.124	0.053	0.048	0.079	0.069	0.027	0.098	0.042	0.029	0.078	0.034	0.025

Table 6: Coverage probabilities for uniform distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.912	0.898	0.910	0.892	0.890	0.906	0.998	1.000	1.000	1.000	1.000	1.000
0.950	0.958	0.949	0.952	0.943	0.946	0.955	1.000	1.000	1.000	1.000	1.000	1.000
0.990	0.988	0.989	0.987	0.991	0.990	0.991	1.000	1.000	1.000	1.000	1.000	1.000

Table 7: Mean and standard deviation of CI lengths for uniform distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.079	0.051	0.038	0.068	0.045	0.033	0.068	0.044	0.032	0.053	0.036	0.026
	0.012	0.006	0.004	0.009	0.005	0.003	0.044	0.028	0.022	0.037	0.024	0.018
0.950	0.106	0.069	0.052	0.092	0.061	0.045	0.077	0.049	0.036	0.059	0.039	0.029
	0.016	0.008	0.005	0.012	0.007	0.004	0.049	0.030	0.023	0.039	0.026	0.020
0.990	0.173	0.113	0.085	0.148	0.098	0.073	0.097	0.059	0.043	0.071	0.047	0.035
	0.027	0.015	0.010	0.021	0.012	0.008	0.055	0.034	0.026	0.043	0.029	0.022

Table 8: Coverage probabilities for skew normal distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.902	0.912	0.902	0.893	0.915	0.902	0.978	0.990	0.978	0.984	0.992	0.989
0.950	0.949	0.956	0.946	0.948	0.957	0.945	0.988	0.994	0.991	0.990	0.995	0.994
0.990	0.988	0.990	0.987	0.992	0.992	0.987	0.993	0.998	0.996	0.997	0.998	0.999

Table 9: Mean and standard deviation of CI lengths for skew normal distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.080	0.053	0.040	0.071	0.046	0.034	0.070	0.046	0.035	0.058	0.038	0.030
	0.015	0.008	0.005	0.013	0.006	0.004	0.046	0.030	0.022	0.039	0.025	0.021
0.950	0.110	0.073	0.054	0.097	0.063	0.047	0.079	0.052	0.039	0.065	0.042	0.033
	0.021	0.012	0.008	0.018	0.009	0.006	0.051	0.033	0.023	0.042	0.028	0.022
0.990	0.182	0.121	0.090	0.160	0.104	0.077	0.099	0.064	0.048	0.079	0.051	0.040
	0.038	0.021	0.014	0.032	0.016	0.011	0.060	0.039	0.027	0.047	0.032	0.025

Table 10: Coverage probabilities for skew t distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.909	0.902	0.910	0.910	0.907	0.902	0.952	0.957	0.951	0.954	0.970	0.959
0.950	0.951	0.958	0.955	0.951	0.945	0.943	0.960	0.970	0.963	0.962	0.973	0.968
0.990	0.986	0.990	0.989	0.987	0.978	0.983	0.979	0.978	0.982	0.981	0.980	0.985

Table 11: Mean and standard deviation of CI lengths for skew t distribution for model (S.3.2)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.076	0.048	0.036	0.063	0.041	0.030	0.067	0.043	0.032	0.053	0.034	0.025
	0.035	0.012	0.009	0.022	0.010	0.006	0.044	0.028	0.022	0.037	0.025	0.017
0.950	0.106	0.067	0.051	0.087	0.056	0.042	0.076	0.048	0.036	0.058	0.038	0.028
	0.050	0.019	0.014	0.031	0.015	0.009	0.049	0.030	0.024	0.039	0.026	0.018
0.990	0.180	0.115	0.087	0.147	0.095	0.071	0.094	0.059	0.045	0.070	0.045	0.033
	0.083	0.036	0.027	0.055	0.027	0.018	0.058	0.035	0.027	0.044	0.029	0.020

Table 12: Coverage probabilities for normal distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.798	0.842	0.856	0.828	0.858	0.882	0.942	0.954	0.955	0.981	0.989	0.992
0.950	0.864	0.915	0.912	0.884	0.904	0.922	0.968	0.974	0.973	0.992	0.996	0.996
0.990	0.944	0.971	0.970	0.946	0.968	0.973	0.990	0.993	0.994	1.000	0.999	1.000

Table 13: Mean and standard deviation of CI lengths for normal distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.313	0.185	0.134	0.253	0.152	0.106	0.416	0.240	0.169	0.303	0.177	0.123
	0.203	0.102	0.068	0.179	0.090	0.050	0.276	0.141	0.107	0.229	0.108	0.064
0.950	0.410	0.251	0.183	0.334	0.206	0.147	0.498	0.293	0.200	0.359	0.207	0.142
	0.221	0.118	0.080	0.198	0.104	0.063	0.321	0.171	0.121	0.261	0.118	0.070
0.990	0.622	0.405	0.298	0.518	0.335	0.245	0.689	0.398	0.269	0.497	0.277	0.189
	0.246	0.144	0.098	0.226	0.130	0.085	0.400	0.209	0.144	0.334	0.153	0.088

Table 14: Coverage probabilities for t distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.802	0.832	0.849	0.801	0.836	0.842	0.917	0.913	0.927	0.958	0.947	0.956
0.950	0.851	0.891	0.903	0.866	0.893	0.906	0.943	0.947	0.948	0.974	0.963	0.970
0.990	0.911	0.954	0.960	0.932	0.946	0.955	0.973	0.969	0.974	0.986	0.979	0.985

Table 15: Mean and standard deviation of CI lengths for t distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.309	0.192	0.133	0.310	0.186	0.129	0.316	0.189	0.134	0.285	0.182	0.136
	0.335	0.209	0.143	0.368	0.206	0.130	0.223	0.121	0.098	0.199	0.134	0.103
0.950	0.408	0.256	0.178	0.402	0.247	0.176	0.373	0.218	0.156	0.333	0.212	0.155
	0.381	0.238	0.160	0.408	0.232	0.153	0.252	0.141	0.110	0.222	0.150	0.110
0.990	0.631	0.402	0.287	0.611	0.392	0.285	0.500	0.288	0.206	0.450	0.283	0.201
	0.459	0.285	0.191	0.474	0.280	0.192	0.306	0.174	0.133	0.282	0.188	0.126

Table 16: Coverage probabilities for uniform distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.770	0.853	0.841	0.805	0.860	0.872	0.943	0.958	0.959	0.979	0.991	0.992
0.950	0.845	0.903	0.902	0.872	0.917	0.933	0.966	0.982	0.975	0.994	0.996	0.999
0.990	0.918	0.957	0.959	0.938	0.966	0.976	0.993	0.998	0.997	1.000	0.999	0.999

Table 17: Mean and standard deviation of CI lengths for uniform distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.367	0.225	0.161	0.285	0.169	0.114	0.565	0.331	0.225	0.389	0.214	0.146
	0.210	0.118	0.078	0.185	0.098	0.053	0.366	0.216	0.148	0.312	0.153	0.090
0.950	0.474	0.299	0.217	0.375	0.227	0.157	0.671	0.396	0.267	0.460	0.252	0.171
	0.222	0.132	0.088	0.206	0.113	0.065	0.406	0.247	0.164	0.346	0.170	0.100
0.990	0.706	0.464	0.345	0.572	0.362	0.260	0.906	0.543	0.367	0.638	0.343	0.227
	0.236	0.147	0.103	0.234	0.136	0.087	0.474	0.301	0.205	0.442	0.225	0.127

Table 18: Coverage probabilities for skew normal distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.809	0.833	0.857	0.847	0.855	0.846	0.947	0.941	0.942	0.982	0.978	0.985
0.950	0.880	0.892	0.919	0.914	0.918	0.903	0.981	0.966	0.970	0.993	0.991	0.992
0.990	0.935	0.956	0.979	0.961	0.971	0.973	0.994	0.994	0.988	0.998	0.999	0.998

Table 19: Mean and standard deviation of CI lengths for skew normal distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.302	0.183	0.127	0.215	0.135	0.095	0.399	0.232	0.167	0.266	0.150	0.114
	0.201	0.107	0.065	0.161	0.086	0.049	0.262	0.150	0.101	0.199	0.086	0.062
0.950	0.395	0.246	0.174	0.287	0.184	0.132	0.474	0.276	0.198	0.311	0.176	0.133
	0.221	0.121	0.077	0.184	0.100	0.060	0.291	0.170	0.111	0.229	0.102	0.069
0.990	0.596	0.391	0.286	0.453	0.301	0.219	0.653	0.376	0.267	0.421	0.236	0.173
	0.244	0.144	0.097	0.224	0.127	0.083	0.361	0.211	0.144	0.289	0.130	0.085

Table 20: Coverage probabilities for skew t distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing I			diminishing II		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.791	0.831	0.832	0.774	0.825	0.835	0.924	0.929	0.925	0.964	0.962	0.961
0.950	0.845	0.881	0.891	0.823	0.877	0.893	0.947	0.949	0.956	0.974	0.977	0.977
0.990	0.923	0.948	0.960	0.912	0.949	0.962	0.979	0.980	0.977	0.994	0.989	0.986

Table 21: Mean and standard deviation of CI lengths for skew t distribution for model (S.3.3)

	fixed r_0			fixed s_0			diminishing II r_0			diminishing II s_0		
$1-\alpha$	600	900	1200	600	900	1200	600	900	1200	600	900	1200
0.900	0.355	0.210	0.141	0.368	0.216	0.137	0.379	0.220	0.162	0.378	0.223	0.153
	0.390	0.201	0.125	0.408	0.229	0.130	0.255	0.149	0.094	0.297	0.166	0.093
0.950	0.455	0.280	0.193	0.468	0.283	0.186	0.451	0.261	0.189	0.449	0.263	0.177
	0.421	0.227	0.149	0.439	0.255	0.150	0.290	0.178	0.107	0.331	0.184	0.101
0.990	0.687	0.446	0.314	0.696	0.441	0.302	0.612	0.347	0.248	0.618	0.349	0.234
	0.477	0.277	0.191	0.490	0.302	0.186	0.364	0.212	0.135	0.412	0.227	0.128

S.4 Proofs of Theorems

S.4.1 Proof of Theorem 3.1

Before the proof of Theorem 3.1, we first give two lemmas. Let $e_t(\boldsymbol{\theta}) = y_t - \sum_{j=1}^4 \boldsymbol{\beta}_j' \mathbf{y}_{t-1} I_{jt}(r,s)$. Recall that e_t is just the $e_t(\boldsymbol{\theta})$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Lemma S.4.1. *If the conditions in Theorem 3.1 hold, then $\mathbb{E}e_t^2(\boldsymbol{\theta}) \geq \mathbb{E}e_t^2$ for all $\boldsymbol{\theta} \in \Theta$, and the equality holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Proof. Because ϵ_t and $\mathcal{F}_{t-1}$ are independent, a conditional argument yields that

$$\begin{aligned}\mathbb{E}e_t^2(\boldsymbol{\theta}) &= \mathbb{E}[e_t(\boldsymbol{\theta}) - e_t]^2 + 2\mathbb{E}\{e_t[e_t(\boldsymbol{\theta}) - e_t]\} + \mathbb{E}e_t^2 \\ &= \mathbb{E}[e_t(\boldsymbol{\theta}) - e_t]^2 + \mathbb{E}e_t^2 \geq \mathbb{E}e_t^2,\end{aligned}$$

for all $\boldsymbol{\theta} \in \Theta$. The equality holds if and only if $\mathbb{E}[e_t(\boldsymbol{\theta}^*) - e_t]^2 = 0$ for some $\boldsymbol{\theta}^* \in \Theta$, which is equivalent to that $e_t(\boldsymbol{\theta}^*) - e_t = 0$ a.s. Since

$$\begin{aligned}e_t(\boldsymbol{\theta}^*) - e_t &= \sum_{j=1}^4 \boldsymbol{\beta}_{j0}' \mathbf{y}_{t-1} I_{jt}(\boldsymbol{\tau}_0) - \sum_{i=1}^4 \boldsymbol{\beta}_i^{*\prime} \mathbf{y}_{t-1} I_{it}(\boldsymbol{\tau}^*) \\ &= \Big[\sum_{j=1}^4 \sum_{i=1}^4 (\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_i^*)' I_{jt}(\boldsymbol{\tau}_0) I_{it}(\boldsymbol{\tau}^*)\Big] \mathbf{y}_{t-1},\end{aligned}$$

we have

$$\sum_{j=1}^4 \sum_{i=1}^4 (\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_i^*) I_{jt}(\boldsymbol{\tau}_0) I_{it}(\boldsymbol{\tau}^*) = 0 \quad \text{a.s.}.$$

By the orthogonality among the indicator functions above, it follows that

$$\sum_{j=1}^4 \sum_{i=1}^4 \|\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_i^*\| \mathbb{E}[I_{jt}(\boldsymbol{\tau}_0) I_{it}(\boldsymbol{\tau}^*)] = 0. \tag{S.4.1}$$

We first prove $r^* = r_0$. Suppose $r^* < r_0$. If $s^* < s_0$, by Assumption 3.3, it follows that

$$\begin{aligned}
&\mathbb{E}[I_{1t}(\boldsymbol{\tau}_0)I_{1t}(\boldsymbol{\tau}^*)] = \mathbb{P}(z_{t-1} > r_0, w_{t-1} > s_0) > 0,\\
&\mathbb{E}[I_{2t}(\boldsymbol{\tau}_0)I_{1t}(\boldsymbol{\tau}^*)] = \mathbb{P}(r^* < z_{t-1} \le r_0, w_{t-1} > s_0) > 0,\\
&\mathbb{E}[I_{3t}(\boldsymbol{\tau}_0)I_{1t}(\boldsymbol{\tau}^*)] = \mathbb{P}(r^* < z_{t-1} \le r_0, s^* < w_{t-1} \le s_0) > 0,\\
&\mathbb{E}[I_{4t}(\boldsymbol{\tau}_0)I_{1t}(\boldsymbol{\tau}^*)] = \mathbb{P}(z_{t-1} > r_0, s^* < w_{t-1} \le s_0) > 0.
\end{aligned}$$

Then, by (S.4.1), we have $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20} = \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_{40}$. Similarly, if $s^* \ge s_0$, then $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20}$ and $\boldsymbol{\beta}_4^* = \boldsymbol{\beta}_{30} = \boldsymbol{\beta}_{40}$. Both cases contradict condition (ii) in Theorem 3.1. Thus, $r^* \ge r_0$. A similar argument can show $r^* \le r_0$ and in turn $r^* = r_0$. Second, we can similarly prove $s^* = s_0$. Finally, by the orthogonality among the indicator functions again, it follows that $\boldsymbol{\beta}_j^* = \boldsymbol{\beta}_{j0}, j = 1, ..., 4$. Therefore $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$. □

Lemma S.4.2. *For any $\boldsymbol{\theta} \in \Theta$ and any $\eta > 0$, define an open neighborhood of $\boldsymbol{\theta}$ as $U_{\boldsymbol{\theta}}(\eta) = \{\boldsymbol{\theta}^* \in \Theta : \|\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_j\| < \eta, |r^* - r| < \eta, |s^* - s| < \eta, j = 1, ..., 4\}$. If the conditions in Theorem 3.1 hold, then*

$$\mathbb{E} \sup_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}}(\eta)} |e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta})| \to 0 \quad \textit{as} \quad \eta \to 0.$$

Proof. Clearly, $e_t^2(\boldsymbol{\theta})$ can be decomposed as follows

$$\begin{aligned}
e_t^2(\boldsymbol{\theta}) &= e_t^2 + 2e_t[e_t(\boldsymbol{\theta}) - e_t] + [e_t(\boldsymbol{\theta}) - e_t]^2\\
&= \sum_{j=1}^4 \sum_{i=1}^4 \{e_t^2 + 2e_t(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_i)'\mathbf{y}_{t-1} + [(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_i)'\mathbf{y}_{t-1}]^2\} I_{jt}(\boldsymbol{\tau}_0) I_{it}(\boldsymbol{\tau})\\
&\equiv \sum_{j=1}^4 \sum_{i=1}^4 [\phi_{ij1}(\boldsymbol{\theta}) + \phi_{ij2}(\boldsymbol{\theta}) + \phi_{ij3}(\boldsymbol{\theta})].
\end{aligned}$$

It suffices to prove for each $\phi_{ijk}(\boldsymbol{\theta})$, $i, j = 1, 2, 3, 4$, $k = 1, 2, 3$,

$$\mathbb{E} \sup_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}}} |\phi_{ijk}(\boldsymbol{\theta}^*) - \phi_{ijk}(\boldsymbol{\theta})| \to 0 \quad \text{as} \quad \eta \to 0. \tag{S.4.2}$$

For example, for $\phi_{11k}(\boldsymbol{\theta})$, $k = 1, 2, 3$, we have

$$\begin{aligned}
|\phi_{111}(\boldsymbol{\theta}^*) - \phi_{111}(\boldsymbol{\theta})| =& e_t^2 I_{1t}(\boldsymbol{\tau}_0) \big| I_{1t}(\boldsymbol{\tau}^*) - I_{1t}(\boldsymbol{\tau}) \big| \le e_t^2 I(\{|z_{t-1} - r| \le \eta\} \cup \{|w_{t-1} - s| \le \eta\}),\\
|\phi_{112}(\boldsymbol{\theta}^*) - \phi_{112}(\boldsymbol{\theta})| =& \big| e_t(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1^*)'\mathbf{y}_{t-1} I_{1t}(\boldsymbol{\tau}_0) I_{1t}(\boldsymbol{\tau}^*) - e_t(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1)'\mathbf{y}_{t-1} I_{1t}(\boldsymbol{\tau}_0) I_{1t}(\boldsymbol{\tau}) \big|\\
\le& \eta \|e_t \mathbf{y}_{t-1}\| + 2\|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1\| \|e_t \mathbf{y}_{t-1}\| I(\{|z_{t-1} - r| \le \eta\} \cup \{|w_{t-1} - s| \le \eta\}),\\
|\phi_{113}(\boldsymbol{\theta}^*) - \phi_{113}(\boldsymbol{\theta})| =& \big| [(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1^*)'\mathbf{y}_{t-1}]^2 I_{1t}(\boldsymbol{\tau}_0) I_{1t}(\boldsymbol{\tau}^*) - [(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1)'\mathbf{y}_{t-1}]^2 I_{1t}(\boldsymbol{\tau}_0) I_{1t}(\boldsymbol{\tau}) \big|\\
\le& 2\eta(\|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1\| + \eta) \|\mathbf{y}_{t-1}\|^2\\
&+ \|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1\|^2 \|\mathbf{y}_{t-1}\|^2 I(\{|z_{t-1} - r| \le \eta\} \cup \{|w_{t-1} - s| \le \eta\}).
\end{aligned}$$

Thus, (S.4.2) holds for $\phi_{11k}(\boldsymbol{\theta})$, $k = 1, 2, 3$, by $\mathbb{E}y_t^2 < \infty$ and the dominated convergence theorem. The other cases can be proved similarly. □

We adopt the approach in Huber (1967) to complete the proof. For any given open neighborhood V of $\boldsymbol{\theta}_0 \in \Theta$ and any $\boldsymbol{\theta} \in V^c = \Theta \backslash V$, it follows that $\mathbb{E}e_t^2(\boldsymbol{\theta}) > \mathbb{E}e_t^2(\boldsymbol{\theta}_0)$ by Lemma S.4.1. Lemma S.4.2 implies that $\mathbb{E}e_t^2(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Applying the compactness of V^c, there exists a $\kappa > 0$ such that

$$\inf_{\boldsymbol{\theta} \in V^c} \mathbb{E}e_t^2(\boldsymbol{\theta}) - \mathbb{E}e_t^2(\boldsymbol{\theta}_0) \geq 3\kappa > 0.$$

For any $\boldsymbol{\theta} \in V^c$, by Lemma S.4.2 again, there exists an $\eta > 0$ such that

$$\mathbb{E} \inf_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}}(\eta)} [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] \geq \mathbb{E}e_t^2(\boldsymbol{\theta}) - \kappa - \mathbb{E}e_t^2(\boldsymbol{\theta}_0) \geq 2\kappa.$$

Since V^c is compact, there exists a finite covering of V^c: $\{U_{\boldsymbol{\theta}_j}(\eta), \boldsymbol{\theta}_j \in V^c, j = 1, 2, ..., T\}$ such that $V^c \subset \cup_{j=1}^T U_{\boldsymbol{\theta}_j}(\eta)$. Since y_t is stationary and ergodic, by the strong law of large numbers, we have a.s.

$$\begin{aligned}\inf_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}_j}(\eta)} \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] &\geq \frac{1}{n} \sum_{t=1}^n \inf_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}_j}(\eta)} [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] \\ &\geq \mathbb{E} \inf_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}}(\eta)} [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] - \kappa \geq \kappa\end{aligned}$$

for n large enough and each $1 \leq j \leq T$. Note that

$$\inf_{\boldsymbol{\theta} \in V} \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}) - e_t^2(\boldsymbol{\theta}_0)] \leq \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}_0) - e_t^2(\boldsymbol{\theta}_0)] = 0.$$

Thus, for any neighborhood V of $\boldsymbol{\theta}_0$, it follows that for n large enough,

$$\begin{aligned}\inf_{\boldsymbol{\theta}^* \in V^c} \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] &= \min_{1 \leq j \leq T} \inf_{\boldsymbol{\theta}^* \in U_{\boldsymbol{\theta}_j}(\eta)} \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}^*) - e_t^2(\boldsymbol{\theta}_0)] \\ &\geq \kappa > 0 \geq \inf_{\boldsymbol{\theta} \in V} \frac{1}{n} \sum_{t=1}^n [e_t^2(\boldsymbol{\theta}) - e_t^2(\boldsymbol{\theta}_0)],\end{aligned}$$

which implies that $\widehat{\boldsymbol{\theta}}_n \in V$ a.s. By the arbitrariness of V, we have $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s.. The proof is complete.

S.4.2 Proof of Theorem 3.2 (i)

Since $\widehat{\boldsymbol{\theta}}_n$ is consistent, we restrict the parameter space to an open neighborhood of $\boldsymbol{\theta}_0$. Define $V_\delta = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \delta, |r - r_0| < \delta, |s - s_0| < \delta\}$ for some $0 < \delta < 1$ to be determined later.

First consider a simple case where $p = 1$ and there are four regimes. It suffices to prove that for any $\epsilon > 0$, there exists $B > 0$ such that, with probability greater than $1 - \epsilon$,

$$L_n(\boldsymbol{\beta}, \boldsymbol{\tau}) - L_n(\boldsymbol{\beta}, \boldsymbol{\tau}_0) > 0 \quad \text{for } |r - r_0| > B/n,\ |s - s_0| > B/n \text{ and } \boldsymbol{\theta} \in V_\delta.$$

Note that the difference of $e_t^2(\boldsymbol{\beta},\boldsymbol{\tau})$ and $e_t^2(\boldsymbol{\beta},\boldsymbol{\tau}_0)$ is nonzero only in the region $\Lambda_t = I\big(\{r\wedge r_0 < z_{t-1} \le r\vee r_0\}\cup\{s\wedge s_0 < w_{t-1} \le s\vee s_0\}\big)$. Here we only treat the case $r > r_0$, $s > s_0$. Proofs of the other three cases are similar. Write $r = r_0 + u$, $s = s_0 + v$ for some $0 < u,\, v < 1$ and let $\alpha_t^{(j)} = (y_t - \boldsymbol{\beta}_j'\mathbf{y}_{t-1})^2$ and $\alpha_t^{(j0)} = (y_t - \boldsymbol{\beta}_{j0}'\mathbf{y}_{t-1})^2$. Then

$$
\begin{aligned}
L_n(\boldsymbol{\beta},\boldsymbol{\tau}) - L_n(\boldsymbol{\beta},\boldsymbol{\tau}_0) =& \big[L_n(\boldsymbol{\beta}, r, s_0) - L_n(\boldsymbol{\beta},\boldsymbol{\tau}_0)\big] + \big[L_n(\boldsymbol{\beta},\boldsymbol{\tau}) - L_n(\boldsymbol{\beta}, r, s_0)\big]\\
\equiv& R_n(\boldsymbol{\beta}, u) + Q_n(\boldsymbol{\theta}, v),
\end{aligned}
$$

where

$$
\begin{aligned}
R_n(\boldsymbol{\beta}, u) =& \sum_{t=1}^{n}\big[(\alpha_t^{(2)} - \alpha_t^{(1)})I(w_{t-1} > s_0) + (\alpha_t^{(3)} - \alpha_t^{(4)})I(w_{t-1} \le s_0)\big]I(r_0 < z_{t-1} \le r_0 + u),\\
Q_n(\boldsymbol{\theta}, v) =& \sum_{t=1}^{n}\big[(\alpha_t^{(4)} - \alpha_t^{(1)})I(z_{t-1} > r) + (\alpha_t^{(3)} - \alpha_t^{(2)})I(z_{t-1} \le r)\big]I(s_0 < w_{t-1} \le s_0 + v).
\end{aligned}
$$

Thus, it suffices to prove that, for any $\epsilon > 0$, there exist constants $B > 0$ and $\gamma > 0$ such that

$$
\mathbb{P}\left(\inf_{\substack{B/n<u<\delta\\ \boldsymbol{\theta}\in V_\delta}} \frac{R_n(\boldsymbol{\beta}, u)}{nG(u)} > \gamma\right) \ge 1-\epsilon \quad\text{and}\quad \mathbb{P}\left(\inf_{\substack{B/n<v<\delta\\ \boldsymbol{\theta}\in V_\delta}} \frac{Q_n(\boldsymbol{\theta}, v)}{nK(v)} > \gamma\right) \ge 1-\epsilon,
$$

where $G(u) = \mathbb{P}(r_0 < z_{t-1} \le r_0 + u)$ and $K(v) = \mathbb{P}(s_0 < w_{t-1} \le s_0 + v)$.

Consider $R_n(\boldsymbol{\beta}, u)$, which can be decomposed as

$$
\begin{aligned}
R_n(\boldsymbol{\beta}, u) =& \big[L_n(\boldsymbol{\beta}_0, r, s_0) - L_n(\boldsymbol{\beta}_0,\boldsymbol{\tau}_0)\big]\\
&+ \big\{[L_n(\boldsymbol{\beta}, r, s_0) - L_n(\boldsymbol{\beta}_0, r, s_0)] - [L_n(\boldsymbol{\beta},\boldsymbol{\tau}_0) - L_n(\boldsymbol{\beta}_0,\boldsymbol{\tau}_0)]\big\}\\
\equiv& R_n^{(1)}(u) + R_n^{(2)}(\boldsymbol{\beta}, u),
\end{aligned}
\tag{S.4.3}
$$

where

$$
\begin{aligned}
R_n^{(1)}(u) =& \sum_{t=1}^{n}\big[(\alpha_t^{(20)} - \alpha_t^{(10)})I(w_{t-1} > s_0) + (\alpha_t^{(30)} - \alpha_t^{(40)})I(w_{t-1} \le s_0)\big]I(r_0 < z_{t-1} \le r_0 + u)\\
=& \sum_{t=1}^{n}\big[\xi_t^{(1,2)}I(w_{t-1} > s_0) + \xi_t^{(4,3)}I(w_{t-1} \le s_0)\big]I\big(r_0 < z_{t-1} \le r_0 + u\big),\\
R_n^{(2)}(\boldsymbol{\beta}, u) =& \sum_{t=1}^{n}\Big\{\big[(\alpha_t^{(2)} - \alpha_t^{(20)}) - (\alpha_t^{(1)} - \alpha_t^{(10)})\big]I(w_{t-1} > s_0)\\
&+ \big[(\alpha_t^{(3)} - \alpha_t^{(30)}) - (\alpha_t^{(4)} - \alpha_t^{(40)})\big]I(w_{t-1} \le s_0)\Big\}I(r_0 < z_{t-1} \le r_0 + u).
\end{aligned}
$$

For $R_n^{(1)}(u)$, by Assumption 3.6, in the four-regime case, for all j there exist some positive constants c_0 and d such that $|(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}| \ge c_0 > 0$ for all $\mathbf{y}_{t-1}$ satisfying $\|\mathbf{y}_{t-1} - \Gamma\| \le d$. Then

$$
\begin{aligned}
\xi_t^{(i,j)} =& [(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}]^2 + 2e_t(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}\\
\ge& c_0^2 I(\|\mathbf{y}_{t-1} - \Gamma\| \le d) + 2e_t(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}.
\end{aligned}
\tag{S.4.4}
$$

Let $\omega = \max_{1\le i,j\le 4} \|\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0}\|$. By (S.4.4) and Lemma S.5.1, for $R_n^{(1)}(u)$, we have

$$R_n^{(1)}(u) \ge c_0^2 G_n^*(u) - 2\omega\Big(\Big|\sum_{t=1}^n A_t(u)\Big| + \Big|\sum_{t=1}^n D_t(u)\Big|\Big),$$

where $G_n^*(u)$, $A_t(u)$, $D_t(u)$, and other notations used below are defined in Lemma S.5.1. Thus,

$$\begin{aligned}
&\inf_{B/n<u<\delta} \frac{R_n^{(1)}(u)}{nG(u)} \\
\ge& c_0^2 \inf_{B/n<u<\delta} \frac{G_n^*(u)}{nG^*(u)} \frac{G^*(u)}{G(u)} - 2\omega \sup_{B/n<u<\delta} \frac{\big|\sum_{t=1}^n A_t(u)\big| + \big|\sum_{t=1}^n D_t(u)\big|}{nG(u)} \\
\ge& c_0^2 \Big(1 - \sup_{B/n<u<\delta}\Big|\frac{G_n^*(u)}{nG^*(u)} - 1\Big|\Big) \frac{G^*(u)}{G(u)} - 2\omega \sup_{B/n<u<\delta} \frac{\big|\sum_{t=1}^n A_t(u)\big| + \big|\sum_{t=1}^n D_t(u)\big|}{nG(u)}.
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{G^*(u)}{G(u)} =& \mathbb{P}(\|\mathbf{y}_{t-1} - \Gamma\| \le d | r_0 < z_{t-1} \le r_0 + u) \\
\to& \mathbb{P}(\|\mathbf{y}_{t-1} - \Gamma\| \le d | z_{t-1} = r_0) > 0 \quad \text{as} \quad u \downarrow 0,
\end{aligned}$$

which implies that the infimum below exists for sufficiently small $\delta > 0$, namely,

$$c_1 \equiv \inf_{0<u<\delta} \frac{G^*(u)}{G(u)} > 0.$$

Choose $\eta_0 > 0$ such that $2\gamma = c_0^2 c_1 - (c_0^2 c_1 + 4\omega)\eta_0 > 0$. By Lemma S.5.1 (ii)-(iv), it follows that

$$\begin{aligned}
&\mathbb{P}\Big(\inf_{B/n<u<\delta} \frac{R_n^{(1)}(u)}{nG(u)} > 2\gamma\Big) \\
\ge& \mathbb{P}\Big(\sup_{B/n<u<\delta}\Big|\frac{G_n^*(u)}{nG^*(u)} - 1\Big| < \eta_0, \sup_{B/n<u<\delta} \frac{\big|\sum_{t=1}^n A_t(u)\big|}{nG(u)} < \eta_0, \sup_{B/n<u<\delta} \frac{\big|\sum_{t=1}^n D_t(u)\big|}{nG(u)} < \eta_0\Big) \\
\ge& 1 - \epsilon. \qquad \text{(S.4.5)}
\end{aligned}$$

Note that

$$\begin{aligned}
&\big[(\alpha_t^{(2)} - \alpha_t^{(20)}) - (\alpha_t^{(1)} - \alpha_t^{(10)})\big] I(w_{t-1} > s_0) I(r_0 < z_{t-1} \le r_0 + u) \\
&= \big\{\big[(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2)'\mathbf{y}_{t-1}\big]^2 - \big[(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_1)'\mathbf{y}_{t-1}\big]^2 + 2(\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2)'\mathbf{y}_{t-1}(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20})'\mathbf{y}_{t-1} \qquad \text{(S.4.6)} \\
&\quad + 2e_t(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{10} + \boldsymbol{\beta}_{20} - \boldsymbol{\beta}_2)'\mathbf{y}_{t-1}\big\} I(w_{t-1} > s_0) I(r_0 < z_{t-1} \le r_0 + u).
\end{aligned}$$

For $R_n^{(2)}(\boldsymbol{\beta}, u)$, by (S.4.6) and Lemma S.5.1, we have

$$|R_n^{(2)}(\boldsymbol{\beta}, u)| \le C\delta\Big(\mathbb{E}(\|\mathbf{y}_1\|^2) G_n(u) + \Big|\sum_{t=1}^n H_t(u)\Big| + \Big|\sum_{t=1}^n A_t(u)\Big| + \Big|\sum_{t=1}^n D_t(u)\Big|\Big),$$

which leads to

$$\sup_{\substack{B/n<u<\delta \\ \boldsymbol{\theta}\in V_\delta}} \frac{|R_n^{(2)}(\boldsymbol{\beta},u)|}{nG(u)} \leq C\delta\Bigg\{\mathbb{E}(\|\mathbf{y}_1\|^2)\Big[\sup_{B/n<u<\delta}\Big|\frac{G_n(u)}{nG(u)}-1\Big|+1\Big] + \sup_{B/n<u<\delta}\frac{|\sum_{t=1}^n H_t(u)|+|\sum_{t=1}^n A_t(u)|+|\sum_{t=1}^n D_t(u)|}{nG(u)}\Bigg\} = O_p(\delta). \tag{S.4.7}$$

Then, by (S.4.3), (S.4.5) and (S.4.7), for sufficiently small $\delta>0$, we have

$$\begin{aligned}\mathbb{P}\Bigg(\inf_{\substack{B/n<u<\delta \\ \boldsymbol{\theta}\in V_\delta}} \frac{R_n(\boldsymbol{\beta},u)}{nG(u)}>\gamma\Bigg) \geq & \mathbb{P}\Bigg(\inf_{B/n<u<\delta}\frac{R_n^{(1)}(u)}{nG(u)} - \sup_{\substack{B/n<u<\delta \\ \boldsymbol{\theta}\in V_\delta}}\frac{|R_n^{(2)}(\boldsymbol{\beta},u)|}{nG(u)}>\gamma\Bigg)\\ \geq & 1-\mathbb{P}\Big(\inf_{B/n<u<\delta}\frac{R_n^{(1)}(u)}{nG(u)}\leq 2\gamma\Big) - \mathbb{P}\Bigg(\sup_{\substack{B/n<u<\delta \\ \boldsymbol{\theta}\in V_\delta}}\frac{|R_n^{(2)}(\boldsymbol{\beta},u)|}{nG(u)}\geq\gamma\Bigg)\\ \geq & 1-2\epsilon.\end{aligned}$$

As for $Q_n(\boldsymbol{\theta},v)$, using the similar technique, we can obtain the result. Then the proofs for four-regime cases are complete.

Now consider the three-regime and two-regime cases. The only difference is that (S.4.4) is not satisfied for all j. All the indicator functions in Lemma S.5.1(ii), (iii) and (iv) need to be multiplied by $I(w_{t-1}\leq s_0)$. Then, with suitable modifications, the preceding proof would go through. This completes the proof for $p=1$.

For general p, replace $D_t(u)=y_{t-1}e_tI(r_0<z_{t-1}\leq r_0+u)$ by $D_t(u)=y_{t-i}e_tI(r_0<z_{t-1}\leq r_0+u)$, where $i=1,...,p$ and the preceding proof would go through. This finally completes the proof.

S.4.3 Proof of Theorem 3.2 (ii)

Let $l_n(\boldsymbol{\beta},\boldsymbol{\tau})=L_n(\boldsymbol{\beta},\boldsymbol{\tau})/n$. By the Taylor expansion of $\partial l_n(\boldsymbol{\beta},\boldsymbol{\tau})/\partial\boldsymbol{\beta}$, we have

$$0=\frac{\partial l_n(\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}),\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}=\frac{\partial l_n(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}+\frac{\partial^2 l_n(\tilde{\boldsymbol{\beta}},\boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}(\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})-\boldsymbol{\beta}_0), \tag{S.4.8}$$

where $\tilde{\boldsymbol{\beta}}$ lies in between $\widehat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_0$, i.e., $\|\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|\leq\|\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\|$.

Let $\tilde{\Sigma}=\text{diag}(\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4)$. Since $\mathbb{E}(y_t^2)<\infty$, by the law of large numbers, it follows that

$$\frac{\partial^2 l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\to 2\tilde{\Sigma}, \quad \text{a.s. as } n\to\infty.$$

Further, by (S.4.8) and Lemma S.5.2, we have

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\leq B/n}\Big\|\sqrt{n}\big[\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})-\boldsymbol{\beta}_0\big]+(2\tilde{\Sigma})^{-1}\sqrt{n}\frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\Big\|=o_p(1).$$

Thus,

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\le B/n}\sqrt{n}\|\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})-\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}_0)\|\le\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\le B/n}\left\|\sqrt{n}\big[\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau})-\boldsymbol{\beta}_0\big]+(2\tilde{\Sigma})^{-1}\sqrt{n}\frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\|$$
$$+\left\|\sqrt{n}\big[\widehat{\boldsymbol{\beta}}_n(\boldsymbol{\tau}_0)-\boldsymbol{\beta}_0\big]+(2\tilde{\Sigma})^{-1}\sqrt{n}\frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\|$$
$$=o_p(1).$$

Note that

$$\sqrt{n}\frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}=\frac{2}{\sqrt{n}}\sum_{t=1}^n\frac{\partial e_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}e_t$$

and $\{e_t\,\partial e_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\beta}\}$ is a martingale difference sequence in terms of $\{\mathcal{F}_t\}$. By the martingale central limit theorem in Brown (1971), it follows that

$$\sqrt{n}\frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\to_d\mathcal{N}(0,4\mathring{\Sigma}),$$

where $\mathring{\Sigma}=\text{diag}(\sigma_{10}^2\Sigma_1,\sigma_{20}^2\Sigma_2,\sigma_{30}^2\Sigma_3,\sigma_{40}^2\Sigma_4)$. Thus,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)\to_d\mathcal{N}(0,\Sigma^{-1}).$$

The proof is complete.

S.4.4 Proof of Proposition 3.1

For simplicity, we only consider the case $u>0$, $v>0$ and $p=1$. Other cases are similar. Denote $I_t(u/n,v/n)=I\big(r_0<z_{t-1}\le r_0+u/n,s_0<w_{t-1}\le s_0+v/n\big)$. The cross term is

$$\tilde{L}_n(u,v)-\tilde{R}_n(u)-\tilde{Q}_n(v)=\sum_{t=1}^n\big(\xi_t^{(13)}-\xi_t^{(12)}-\xi_t^{(14)}\big)I_t(u,v). \tag{S.4.9}$$

Recalling $\omega=\max_{1\le i,j\le 4}\|\boldsymbol{\beta}_{i0}-\boldsymbol{\beta}_{j0}\|$, it follows that

$$\xi_t^{(i,j)}=[(\boldsymbol{\beta}_{i0}-\boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}]^2+2e_t(\boldsymbol{\beta}_{i0}-\boldsymbol{\beta}_{j0})'\mathbf{y}_{t-1}\le\omega^2\|\mathbf{y}_{t-1}\|^2+2\omega(|e_t|+|e_ty_{t-1}|).$$

Define

$$G(u/n,v/n)=\mathbb{E}(I_t(u/n,v/n)),\quad A_t(u/n,v/n)=e_tI_t(u/n,v/n),$$
$$D_t(u/n,v/n)=y_{t-1}e_tI_t(u/n,v/n),\quad H_t(u/n,v/n)=(\|\mathbf{y}_{t-1}\|^2-\mathbb{E}\|\mathbf{y}_{t-1}\|^2)I_t(u/n,v/n).$$

Then, there exists a constant $H>0$ such that

$$\sum_{t=1}^n\xi_t^{(i,j)}I_t(u/n,v/n)$$
$$\le H\left\{\sum_{t=1}^nI_t(u/n,v/n)+\left|\sum_{t=1}^nH_t(u/n,v/n)\right|+\left|\sum_{t=1}^nA_t(u/n,v/n)\right|+\left|\sum_{t=1}^nD_t(u/n,v/n)\right|\right\}.$$

It suffices to prove that $\sum_{t=1}^{n} I_t(u/n, v/n)$, $\big|\sum_{t=1}^{n} H_t(u/n, v/n)\big|$, $\big|\sum_{t=1}^{n} A_t(u/n, v/n)\big|$ and $\big|\sum_{t=1}^{n} D_t(u/n, v/n)\big|$ are all $o_p(1)$. And the proof arguments are similar to those of (i), (iii), (iv) and (v) in Lemma S.5.1. We only detail the proof of $\sum_{t=1}^{n} I_t(u/n, v/n)$ and others are similar.

Similar to (S.5.1), (S.5.2) and (S.5.3), by choosing B sufficiently small, there exist $0 < m < M < \infty$ and $H > 0$, independent of n, such that for any $u, v \in [0, B)$,

$$
\begin{aligned}
& muv/n^2 \le G(u/n, v/n) \le Muv/n^2, \\
& \mathrm{Var}(I_t(u/n, v/n)) \le HG(u/n, v/n), \\
& \mathrm{Var}\Big(\sum_{t=1}^{n} I_t(u/n, v/n)\Big) \le nHG(u/n, v/n).
\end{aligned}
$$

Without loss of generality, let $G(u, v) = uv$. Using the Markov inequality, for each $\epsilon > 0$,

$$
\mathbb{P}\left(\sum_{t=1}^{n} I_t(u/n, v/n) > \epsilon\right) \le nHuv/(n^2\epsilon^2) = Huv/(n\epsilon^2) \to 0, \quad \text{as} \quad n \to \infty,
$$

which means $\sum_{t=1}^{n} I_t(u/n, v/n) = o_p(1)$. The other three term are all $o_p(1)$ by similar approach, and thus so is $\sum_{t=1}^{n} \xi_t^{(i,j)} I_t(u/n, v/n)$. By (S.4.9), the result holds.

S.4.5 Proof of Theorem 3.3

Define $\psi_n(u, v) = \tilde{R}_n(u) + \tilde{Q}_n(v)$. First, we show that $\psi_n(u, v)$ converges weakly to a spatial compound Poisson process. The proof consists of two steps: (i) verifying the tightness of $\psi_n(u, v)$; (ii) characterizing the convergence of finite-dimensional distributions.

(i). *Tightness of $\psi_n(u, v)$.* Similar to Li and Ling (2012) and Li et al. (2013), it is not hard to prove the tightness of $\tilde{R}_n(u)$ and $\tilde{Q}_n(v)$, respectively. Thus, $\psi_n(u, v)$ is tight.

(ii). *Convergence of finite-dimensional distributions.* Without loss of generality, we assume $\xi_t^{(i,j)}$ is bounded. Otherwise, use the truncating technique in Li et al. (2013) to truncate $\xi_t^{(i,j)}$ and then consider a truncated process. Here, we only consider the case $u > 0$, $v > 0$. Other cases are similar. For any $0 \le u_1 \le u_2 \le u_3 \le u_4 < \infty$, $0 \le v_1 \le v_2 \le v_3 \le v_4 < \infty$ and any constants c_1 and c_2, the linear combination of the increments of $\psi_n(u, v)$ is

$$
\begin{aligned}
S_n \equiv & c_1\big[\psi_n(u_2, v_2) - \psi_n(u_1, v_1)\big] + c_2\big[\psi_n(u_4, v_4) - \psi_n(u_3, v_3)\big] \\
= & \sum_{t=1}^{n} \Big\{ \big[\xi_t^{(1,2)} I(w_{t-1} > s_0 + v_1/n) + \xi_t^{(4,3)} I(w_{t-1} \le s_0 + v_1/n)\big] c_1 I_{1t}^{(1)} \\
& + \big[\xi_t^{(1,2)} I(w_{t-1} > s_0 + v_3/n) + \xi_t^{(4,3)} I(w_{t-1} \le s_0 + v_3/n)\big] c_2 I_{1t}^{(3)} \\
& + \big[\xi_t^{(1,4)} I(z_{t-1} > r_0 + u_2/n) + \xi_t^{(2,3)} I(z_{t-1} \le r_0 + u_2/n)\big] c_1 I_{2t}^{(1)} \\
& + \big[\xi_t^{(1,4)} I(z_{t-1} > r_0 + u_4/n) + \xi_t^{(2,3)} I(z_{t-1} \le r_0 + u_4/n)\big] c_2 I_{2t}^{(3)} \Big\},
\end{aligned}
$$

where

$$I_{1t}^{(j)} = I\big(r_0 + \frac{u_j}{n} < z_{t-1} \le r_0 + \frac{u_{j+1}}{n}\big), \quad I_{2t}^{(j)} = I\big(s_0 + \frac{v_j}{n} < w_{t-1} \le s_0 + \frac{v_{j+1}}{n}\big)$$

for $j = 1, 3$. Let $\epsilon = 1/n$ and consider the following process indexed by ϵ:

$$\begin{aligned}
x^\epsilon(t) =& X^\epsilon_{[nt]}, 0 \le t \le 1, \\
X^\epsilon_k =& X^\epsilon_{k-1} + J^\epsilon_k, \quad k \ge 1, \quad \text{with } X^\epsilon_0 = 0, \\
J^\epsilon_k =& \big[\xi_k^{(1,2)} I(w_{k-1} > s_0 + v_1\epsilon) + \xi_k^{(4,3)} I(w_{k-1} \le s_0 + v_1\epsilon)\big] c_1 I_{1k}^{(1)} \\
&+ \big[\xi_k^{(1,2)} I(w_{k-1} > s_0 + v_3\epsilon) + \xi_k^{(4,3)} I(w_{k-1} \le s_0 + v_3\epsilon)\big] c_2 I_{1k}^{(3)} \\
&+ \big[\xi_k^{(1,4)} I(z_{k-1} > r_0 + u_2\epsilon) + \xi_k^{(2,3)} I(z_{k-1} \le r_0 + u_2\epsilon)\big] c_1 I_{2k}^{(1)} \\
&+ \big[\xi_k^{(1,4)} I(z_{k-1} > r_0 + u_4\epsilon) + \xi_k^{(2,3)} I(z_{k-1} \le r_0 + u_4\epsilon)\big] c_2 I_{2k}^{(3)}.
\end{aligned}$$

Clearly, $x^\epsilon(1) = S_n$. We now verify Assumptions A.1-A.4 in Li et al. (2013) for J^ϵ_k. By Assumption 3.5, we have

$$\begin{aligned}
\lambda =& \lim_{\epsilon\to 0} \lim_{m\to\infty} \epsilon^{-1} \mathbb{P}^\epsilon_k (J^\epsilon_m \ne 0) \\
=& \lim_{\epsilon\to 0} \lim_{m\to\infty} \epsilon^{-1} \mathbb{E}^\epsilon_k \big[\big(I_{1m}^{(1)} + I_{2m}^{(1)} - I_{1m}^{(1)} I_{2m}^{(1)}\big) + \big(I_{1m}^{(3)} + I_{2m}^{(3)} - I_{1m}^{(3)} I_{2m}^{(3)}\big)\big] \\
=& \pi_1(r_0)[(u_2 - u_1) + (u_4 - u_3)] + \pi_2(s_0)[(v_2 - v_1) + (v_4 - v_3)].
\end{aligned} \tag{S.4.10}$$

By the stationarity of y_t and Assumption 3.5, for any Borel set $\mathfrak{B}$, it follows that

$$\mathbb{Q}(\mathfrak{B}) = \sum_{i=1}^{4} \iota_i \mathbb{Q}_i(\mathfrak{B}), \tag{S.4.11}$$

where

$$\begin{aligned}
\iota_1 =& \pi_1(r_0)(u_2 - u_1)/\lambda, \quad \iota_2 = \pi_1(r_0)(u_4 - u_3)/\lambda, \\
\iota_3 =& \pi_2(s_0)(v_2 - v_1)/\lambda, \quad \iota_4 = \pi_2(r_0)(v_4 - v_3)/\lambda,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{Q}_1(\mathfrak{B}) =& \mathbb{P}\big(c_1 \big[\xi_k^{(1,2)} I(w_{k-1} > s_0) + \xi_k^{(4,3)} I(w_{k-1} \le s_0)\big] \in \mathfrak{B} \big| z_{k-1} = r_0\big), \\
\mathbb{Q}_2(\mathfrak{B}) =& \mathbb{P}\big(c_2 \big[\xi_k^{(1,2)} I(w_{k-1} > s_0) + \xi_k^{(4,3)} I(w_{k-1} \le s_0)\big] \in \mathfrak{B} \big| z_{k-1} = r_0\big), \\
\mathbb{Q}_3(\mathfrak{B}) =& \mathbb{P}\big(c_1 \big[\xi_k^{(1,4)} I(z_{k-1} > r_0) + \xi_k^{(2,3)} I(z_{k-1} \le r_0)\big] \in \mathfrak{B} \big| w_{k-1} = s_0\big), \\
\mathbb{Q}_4(\mathfrak{B}) =& \mathbb{P}\big(c_2 \big[\xi_k^{(1,4)} I(z_{k-1} > r_0) + \xi_k^{(2,3)} I(z_{k-1} \le r_0)\big] \in \mathfrak{B} \big| w_{k-1} = s_0\big).
\end{aligned}$$

Similarly, we can verify that, for any $f \in \widehat{C}_0^2$, a space of functions with compact support and continuous second derivative, and a scalar x,

$$\lim_{\epsilon\to 0} \lim_{t\to\infty} \mathbb{E}^\epsilon_k \big[f(x + J^\epsilon_t) - f(x) | J^\epsilon_t \ne 0\big] = \int [f(x+z) - f(x)] \mathbb{Q}(dz). \tag{S.4.12}$$

By (S.4.10)-(S.4.12) and Theorem A.1 in Li et al. (2013), $x^{\epsilon}(t)$ converges weakly to a compound Poisson process $J(t)$ with jump rate λ and jump distribution $\mathbb{Q}$. Thus, S_n converges weakly to $J(1)$. The characteristic function $f_J(t)$ of $J(1)$ can be written as

$$f_J(t) = \exp\Big\{\lambda\Big[\int_{\mathbb{R}} e^{itx}\mathbb{Q}(dx) - 1\Big]\Big\} = \prod_{i=1}^{4}\exp\Big\{\lambda\iota_i\Big[\int_{\mathbb{R}} e^{itx}\mathbb{Q}_i(dx) - 1\Big]\Big\},$$

which equals that of $c_1\big\{\psi(u_2, v_2) - \psi(u_1, v_1)\big\} + c_2\big\{\psi(u_4, v_4) - \psi(u_3, v_3)\big\}$, the linear combination of the increments of a spatial compound Poisson process

$$\psi(u, v) = \sum_{k=1}^{N_1(u)} \zeta_k^{(1)} + \sum_{k=1}^{N_3(v)} \zeta_k^{(3)}, \quad u \geq 0,\ v \geq 0.$$

Here, $N_1(u)$, $N_3(v)$, $\zeta_k^{(1)}$, and $\zeta_k^{(3)}$ are defined in (3.3) and (3.4), respectively.

Now we will show that any linear combination of $\psi_n(u^{(1)}, v^{(1)})$ and $\psi_n(u^{(2)}, v^{(2)})$ can be written as linear combination of increments. Assume $u > 0, v > 0$. Without loss of generality, assume $0 < u^{(1)} < u^{(2)}$. First consider the case that $0 < v^{(1)} < v^{(2)}$. For any constants $c^{(1)}$ and $c^{(2)}$, we have

$$\begin{aligned}
&c^{(1)}\psi_n(u^{(1)}, v^{(1)}) + c^{(2)}\psi_n(u^{(2)}, v^{(2)})\\
=&c^{(1)}\psi_n(u^{(1)}, v^{(1)}) + c^{(2)}[\psi_n(u^{(2)}, v^{(2)}) - \psi_n(u^{(1)}, v^{(1)}) + \psi_n(u^{(1)}, v^{(1)})]\\
=&(c^{(1)} + c^{(2)})[\psi_n(u^{(1)}, v^{(1)}) - \psi_n(0, 0)] + c^{(2)}[\psi_n(u^{(2)}, v^{(2)}) - \psi_n(u^{(1)}, v^{(1)})].
\end{aligned}$$

Compared with S_n, we have the according relationship: $c_1 = c^{(1)} + c^{(2)}, c_2 = c^{(2)}, u_1 = v_1 = 0, u_2 = u_3 = u^{(1)}, v_2 = v_3 = v^{(1)}, u_4 = u^{(2)}, v_4 = v^{(2)}$ with $0 \leq u_1 \leq u_2 \leq u_3 \leq u_4 < \infty$ and $0 \leq v_1 \leq v_2 \leq v_3 \leq v_4 < \infty$. The order-distortion case $v^{(1)} > v^{(2)}$ is a little different in that

$$\begin{aligned}
&c^{(1)}\psi_n(u^{(1)}, v^{(1)}) + c^{(2)}\psi_n(u^{(2)}, v^{(2)})\\
=&c^{(1)}[\psi_n(u^{(1)}, v^{(1)}) - \psi_n(u^{(1)}, v^{(2)}) + \psi_n(u^{(1)}, v^{(2)})]\\
&+ c^{(2)}[\psi_n(u^{(2)}, v^{(2)}) - \psi_n(u^{(1)}, v^{(2)}) + \psi_n(u^{(1)}, v^{(2)})]\\
=&c^{(1)}[\psi_n(u^{(1)}, v^{(1)}) - \psi_n(u^{(1)}, v^{(2)})] + c^{(2)}[\psi_n(u^{(2)}, v^{(2)}) - \psi_n(u^{(1)}, v^{(2)})]\\
&+ (c^{(1)} + c^{(2)})[\psi_n(u^{(1)}, v^{(2)}) - \psi_n(0, 0)],
\end{aligned} \tag{S.4.13}$$

which is a linear combination of three increments and the according indexes in the increment form do not follow $0 \leq u_1 \leq u_2 \leq u_3 \leq u_4 < \infty$ and $0 \leq v_1 \leq v_2 \leq v_3 \leq v_4 < \infty$. However, noting that the essential reason for S_n's convergence is that the increments in the combination are asymptotically independent in the sense of (S.4.10), which also holds here for the combination in (S.4.13). Thus, following the same procedure for S_n, we could prove the weak convergence of (S.4.13), too.

For the other three cases of u and v, we can obtain a similar result. By the Crámer-Wold device, the finite dimensional distribution of $\psi_n(u, v)$ converges weakly to that of $\psi(u, v)$, i.e., $\psi_n(u, v) \Longrightarrow \psi(u, v)$ in $\mathbb{D}(\mathbb{R}^2)$ as $n \to \infty$. By Theorem 3.1 in Seijo and Sen (2011), it is readily seen that $n(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)$ converges weakly to $\boldsymbol{M}_{-}$. The remainder is similar to the proof of Theorem 2 in Chan (1993).

S.4.6 Proof of Theorem 4.2

In Section 4, Proposition 4.1 follows from Theorem 1 in Cai (2002). The proofs of Theorem 4.1 and Theorem 4.2 follow the same route, the latter being more complicated since it involves estimated parameters. Thus, we only focus on the proof of Theorem 4.2.

Compared to the original resampling algorithm for self-exciting threshold models in Li and Ling (2012), our modified resampling algorithm only differs in the design of $\boldsymbol{Q}$ in order to adapt to the cases of threshold variables being linear functions of lags of y_t. By noting this fact, the proof of Theorem 4.2(ii) is similar to that in Li and Ling (2012) and thus omitted.

We focus on the weighted Nadaraya Watson method in Theorem 4.2(i). To prove (i), we first give a lemma of the $\widetilde{F}_1^{\mathrm{WNW}}(\gamma|\hat{r}_n, \mathcal{X})$ in (S.2.2).

Lemma S.4.3. *Under the conditions in Proposition 4.1, we have in probability that*

$$|\widetilde{F}_1^{\mathrm{WNW}}(\gamma|\hat{r}_n, \mathcal{X}) - F_1(\gamma|r_0)| \to 0.$$

Proof. By a simple algebraic calculation, we have

$$\begin{aligned}
&\widetilde{F}_1^{\mathrm{WNW}}(\gamma|\hat{r}_n, \mathcal{X}) - \widehat{F}_1^{\mathrm{WNW}}(\gamma|r_0, \mathcal{X}) \\
=&\frac{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})} - \frac{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})I(\gamma_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} \\
=&\frac{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})} - \frac{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} \\
+&\frac{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} - \frac{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} \\
+&\frac{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} - \frac{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})I(\gamma_t^{(1)} \le \gamma)}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})} \\
\equiv&J_1 + J_2 + J_3.
\end{aligned}$$

It follows that

$$\begin{aligned}
J_1 =&\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})I(\widehat{\gamma}_t^{(1)} \le \gamma)\left(\frac{1}{\sum_{t=1}^n p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1})} - \frac{1}{\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})}\right) \\
\le&\frac{\frac{1}{n}\sum_{t=1}^n |p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1}) - p_t(r_0)K_h(r_0 - z_{t-1})|}{\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})}, \\
J_2 \le&\frac{\frac{1}{n}\sum_{t=1}^n |p_t(\hat{r}_n)K_h(\hat{r}_n - z_{t-1}) - p_t(r_0)K_h(r_0 - z_{t-1})|}{\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})}, \\
J_3 \le&\frac{\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})|I(\widehat{\gamma}_t^{(1)} \le \gamma) - I(\gamma_t^{(1)} \le \gamma)|}{\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})}.
\end{aligned}$$

By Lemma 2 and Theorem 1 in Cai (2002), we have $\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1}) = \pi_1(r_0) + o_p(1)$, which is positive and bounded by our Assumption 3.3. Since $\hat{r}_n \to r_0$ a.s., using the fact that both $p_t(z)$ and $K_h(z-z_{t-1})$ are continuous function of z, the continuous mapping theorem leads to the fact that $J_1 = o_p(1)$ and $J_2 = o_p(1)$.

Define $\tilde{\gamma}_t^{(1)}$ similar to $\gamma_t^{(1)}$ but with $\boldsymbol{\theta}_0$ replaced by some $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2', \tilde{\boldsymbol{\beta}}_3', \tilde{\boldsymbol{\beta}}_4', \tilde{r}, \tilde{s})'$, and $\boldsymbol{\sigma}$ replaced by some $\tilde{\boldsymbol{\sigma}}$. Specifically,

$$\tilde{\gamma}_t^{(1)} = \tilde{\xi}_t^{(1,2)} I(w_{t-1} > \tilde{s}) + \tilde{\xi}_t^{(4,3)} I(w_{t-1} \leq \tilde{s}),$$

with

$$\tilde{\xi}_t^{(i,j)} = [(\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_j)'\mathbf{y}_{t-1}]^2 + 2\tilde{\sigma}_i \tilde{\epsilon}_t (\tilde{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_j)'\mathbf{y}_{t-1}, \quad i,j = 1,...,4,$$

where $\tilde{\epsilon}_t$ are the corresponding residuals of the 2-TAR model with the parameters $\tilde{\boldsymbol{\theta}}$.

Consider a neighbourhood of $(\boldsymbol{\theta}_0, \boldsymbol{\sigma}_0)$. Define $V_\eta = \{(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}) : \|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq \eta, \|\tilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\| \leq \eta\}$ for some $0 < \eta < 1$. By the definition of $\gamma_t^{(1)}$, we know that it is continuous and $\mathbb{E}\left(\sup_{(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}) \in V_\eta} |\tilde{\gamma}_t^{(1)} - \gamma_t^{(1)}|\right) = O(\eta^2)$. Then, for any $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}) \in V_\eta$ and γ, it follows that

$$\begin{aligned}
\mathbb{E}\big(|I(\tilde{\gamma}_t^{(1)} \leq \gamma) - I(\gamma_t^{(1)} \leq \gamma)|\big) \leq & \mathbb{P}\big(|\gamma_t^{(1)} - \gamma| < |\tilde{\gamma}_t^{(1)} - \gamma_t^{(1)}|\big) \\
\leq & \mathbb{P}\big(|\tilde{\gamma}_t^{(1)} - \gamma_t^{(1)}| > \eta\big) + \mathbb{P}\big(|\gamma_t^{(1)} - \gamma| < |\tilde{\gamma}_t^{(1)} - \gamma_t^{(1)}| \leq \eta\big) \\
\leq & \mathbb{P}\big(|\tilde{\gamma}_t^{(1)} - \gamma_t^{(1)}| > \eta\big) + \mathbb{P}\big(|\gamma_t^{(1)} - \gamma| \leq \eta\big) \\
= & O(\eta).
\end{aligned}$$

By Lemma 2 and Theorem 1 in Cai (2002), we have $\mathbb{E}(p_t(r_0)K_h(r_0 - z_{t-1})) = \pi_1(r_0) + o(1)$. Thus, $\mathbb{E}\big(p_t(r_0)K_h(r_0 - z_{t-1})|I(\tilde{\gamma}_t^{(1)} \leq \gamma) - I(\gamma_t^{(1)} \leq \gamma)|\big) = O(\eta)$. Then, for any $\nu > 0$, noting the consistency of $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\sigma}}_n$, by choosing η small enough, we have

$$\begin{aligned}
& \mathbb{P}\Big(\frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})|I(\widehat{\gamma}_t^{(1)} \leq \gamma) - I(\gamma_t^{(1)} \leq \gamma)| > \nu\Big) \\
\leq & \mathbb{P}\Big(\sup_{(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}) \in V_\eta} \frac{1}{n}\sum_{t=1}^n p_t(r_0)K_h(r_0 - z_{t-1})|I(\tilde{\gamma}_t^{(1)} \leq \gamma) - I(\gamma_t^{(1)} \leq \gamma)| > \nu\Big) + \mathbb{P}((\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\sigma}}) \notin V_\eta) \\
= & o(1).
\end{aligned}$$

Thus, $J_3 = o_p(1)$.

Combining the results, we have in probability that

$$\widetilde{F}_1^{\text{WNW}}(\gamma|\hat{r}_n, \mathcal{X}) - \widehat{F}_1^{\text{WNW}}(\gamma|r_0, \mathcal{X}) \to 0.$$

Since Proposition 4.1 gives that $|\widehat{F}_1^{\text{WNW}}(\gamma|r_0, \mathcal{X}) - F_1(\gamma|r_0)| \to 0$ in probability, we obtain in probability that

$$\widetilde{F}_1^{\text{WNW}}(\gamma|\hat{r}_n, \mathcal{X}) \to F_1(\gamma|r_0).$$

□

Noting that every compound Poisson process is a stationary independent increment process, by the consistency of $\widehat{\pi}_1(\hat{r}_n)$ to $\pi_1(r_0)$, Lemma S.4.3, Theorem 16 in Pollard (1984) and Theorem 3.1 in Seijo and Sen (2011), $\hat{M}_1^{\text{WNW}}$ converges weakly to $M_-^{(1)}$ conditionally on $\mathcal{X}$, in probability. For details of the proof, refer to Li and Ling (2012).

S.5 Proofs of Auxilliary Lemmas

Lemma S.5.1. *If Assumptions 3.1-3.6 hold, then, for any $\epsilon > 0$, $\eta > 0$ and $\delta \in (0,1)$, there exists a positive constant B such that for n large enough, we have*

$$\text{(i).}\quad \mathbb{P}\Big(\sup_{B/n<u<\delta}\Big|\frac{G_n(u)}{nG(u)}-1\Big|<\eta\Big)>1-\epsilon,$$

$$\text{(ii).}\quad \mathbb{P}\Big(\sup_{B/n<u<\delta}\Big|\frac{G_n^*(u)}{nG^*(u)}-1\Big|<\eta\Big)>1-\epsilon,$$

$$\text{(iii).}\quad \mathbb{P}\Big(\sup_{B/n<u<\delta}\frac{|\sum_{t=1}^n A_t(u)|}{nG(u)}<\eta\Big)>1-\epsilon,$$

$$\text{(iv).}\quad \mathbb{P}\Big(\sup_{B/n<u<\delta}\frac{|\sum_{t=1}^n D_t(u)|}{nG(u)}<\eta\Big)>1-\epsilon,$$

$$\text{(v).}\quad \mathbb{P}\Big(\sup_{B/n<u<\delta}\frac{|\sum_{t=1}^n H_t(u)|}{nG(u)}<\eta\Big)>1-\epsilon,$$

where

$$\begin{aligned}
G(u) &= \mathbb{P}(r_0 < z_{t-1} \le r_0+u),\\
G_n(u) &= \sum_{t=1}^n I(r_0 < z_{t-1} \le r_0+u),\\
G_n^*(u) &= \sum_{t=1}^n I(r_0 < z_{t-1} \le r_0+u, \|\mathbf{y}_{t-1}-\Gamma\| \le d),\\
G^*(u) &= \mathbb{P}(r_0 < z_{t-1} \le r_0+u, \|\mathbf{y}_{t-1}-\Gamma\| \le d),\\
A_t(u) &= e_t I(r_0 < z_{t-1} \le r_0+u),\\
D_t(u) &= y_{t-1}e_t I(r_0 < z_{t-1} \le r_0+u),\\
H_t(u) &= (\|\mathbf{y}_{t-1}\|^2 - \mathbb{E}\|\mathbf{y}_{t-1}\|^2) I(r_0 < z_{t-1} \le r_0+u).
\end{aligned}$$

PROOF. We only prove (i), (iv) and (v) since the other two are similar.

(i). By choosing δ sufficiently small, we establish the following inequalities that there exist $0 < m < M < \infty$ and $H > 0$, independent of n, such that any $u \in [0,\delta)$,

$$mu \le G(u) \le Mu, \tag{S.5.1}$$

$$\mathrm{Var}(I(r_0 < z_t \le r_0+u)) \le HG(u), \tag{S.5.2}$$

$$\mathrm{Var}(G_n(u)) \le nHG(u). \tag{S.5.3}$$

Clearly, (S.5.1) is implied by Assumption 3.3, and then (S.5.2) follows. For simplicity, $I_t(u) = I(r_0 < z_t \le r_0+u)$. Recall $\mathbf{x}_t = (y_t, z_t, w_t)'$. By a direct calculation, (S.5.3) is implied by the following fact

$$\begin{aligned}\big|\mathrm{Cov}(I_0(u), I_j(u))\big| =&\big|\mathbb{E}\big(\mathbb{E}\big(I_0(u)I_j(u)|\mathbf{x}_0\big)\big) - \mathbb{E}\big(I_0(u)\big)\mathbb{E}\big(I_j(u)\big)\big| \\ =&\mathbb{E}\big(I_0(u)\big|\mathbb{E}\big(I_j(u)|\mathbf{x}_0\big) - \mathbb{E}\big(I_0(u)\big)\big|\big) \\ \leq&\mathbb{E}\big(I_0(u)\|\mathbb{P}^j(\mathbf{x}_0, \cdot) - \Pi(\cdot)\|_{\mathrm{v}}\big) \\ \leq&C\rho^j\mathbb{E}\big\{(1 + \|\mathbf{x}_0\|)I(r_0 < z_0 \leq r_0 + u)\big\} \\ =&O(\rho^j G(u)),\end{aligned} \tag{S.5.4}$$

which is implied by Assumption 3.5. Without loss of generality, let $G(u) = u$. Then, for a $B > 0$, choose a partition of the region $(B/n, 1]$ as follows: fix a $b > 1$ and let $R_i = (b^i B/n, b^{i+1}B/n]$ for all possible $i \geq 0$. Using the Markov inequality, we obtain

$$\mathbb{P}\Big(\sup_i \Big|\frac{G_n(b^i B/n)}{nG(b^i B/n)} - 1\Big| > \eta\Big) < \sum_{i\geq 0} \frac{H}{nG(b^i B/n)\eta^2} \leq \frac{H}{B\eta^2(1 - b^{-1})}. \tag{S.5.5}$$

For $0 < x \leq y \leq bx \leq \delta$ with $|G_n(x)/(nx) - 1| < \eta$ and $|G_n(bx)/(nbx) - 1| < \eta$, we have

$$\begin{aligned}&G_n(y)/(ny) - 1 \geq G_n(x)/(nbx) - 1 \geq (1 - \eta)/b - 1, \\ &G_n(y)/(ny) - 1 \leq G_n(bx)/(nx) - 1 \leq (1 + \eta)b - 1.\end{aligned} \tag{S.5.6}$$

By choosing $\eta > 0$ and $b > 1$ sufficiently small, and then choosing sufficiently large B, (S.5.5) and (S.5.6) imply the validity of (i).

(iv). Let $\tilde{I}_{t-1}(u_1, u_2) = I(r_0 + u_1 < z_{t-1} \leq r_0 + u_2)$ and

$$\tilde{D}_t(u_1, u_2) = |e_t y_{t-1}|\tilde{I}_{t-1}(u_1, u_2).$$

Then, using the same technique in (S.5.4), we can obtain

$$\begin{aligned}&\Big|\mathrm{Cov}\left(\tilde{D}_1(u_1, u_2), \tilde{D}_j(u_1, u_2)\right)\Big| \\ =&\Big|\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\left[\mathbb{E}\left(|e_j y_{j-1}|\tilde{I}_{j-1}(u_1, u_2)\big|\mathbf{x}_1\right) - \mathbb{E}\left(|e_1 y_0|\tilde{I}_0(u_1, u_2)\right)\right]\right\}\Big| \\ \leq&C\left|\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\left[\mathbb{E}\left(\mathbb{E}\left(|y_{j-1}|\big|z_{j-1}\right)\tilde{I}_{j-1}(u_1, u_2)\big|\mathbf{x}_1\right) - \mathbb{E}\left(\mathbb{E}\left(|y_0|\big|z_0\right)\tilde{I}_0(u_1, u_2)\right)\right]\right\}\right| \\ \leq&C\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\left|\int \mathbb{E}(|y|\big|z)\tilde{I}(u_1, u_2)\left[\mathbb{P}^{j-2}(\mathbf{x}_1, dx) - \Pi(dx)\right]\right|\right\} \\ \leq&C\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\|\mathbb{P}^{j-2}(\mathbf{x}_1, \cdot) - \Pi(\cdot)\|_{\mathrm{v}}\right\} \\ \leq&C\rho^{j-2}\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\left(1 + \|\mathbf{x}_1\|\right)\right\} \\ \leq&C\rho^{j-2}\mathbb{E}\left\{|e_1 y_0|\tilde{I}_0(u_1, u_2)\left(1 + |y_0| + |e_1| + |z_1| + |w_1|\right)\right\} \\ \leq&C\rho^{j-2}\mathbb{E}\left\{\tilde{I}_0(u_1, u_2)\mathbb{E}\left(|e_1 y_0| + |e_1 y_0^2| + |e_1^2 y_0| + |e_1 y_0 z_1| + |e_1 y_0 w_1|\big|z_0\right)\right\} \\ =&O\left(\rho^j\left[G(u_2) - G(u_1)\right]\right),\end{aligned}$$

where Assumptions 3.4 and 3.5 are used such that

$$\mathrm{Var}\Big(\sum_{t=1}^{n} \tilde{D}_t(u_1,u_2)\Big) \le nH\big(G(u_2)-G(u_1)\big). \tag{S.5.7}$$

In addition, we also have

$$\mathrm{Var}\Big(\sum_{t=1}^{n} D_t(u)\Big) \le nHG(u). \tag{S.5.8}$$

Similarly, choose a partition of the region $(B/n,1]$ as follows: for $B>0$, fix $b>1$ and let $R_i=(b^iB/n, b^{i+1}B/n]$ for all possible $i\ge 0$. For $u\in R_i=(b^iB/n, b^{i+1}B/n]$, we can obtain

$$\frac{\sum_{t=1}^{n} D_t(u)}{nG(u)} = \frac{\sum_{t=1}^{n} D_t(b^iB/n)}{nG(u)} + \sum_{t=1}^{n}\frac{D_t(u)-D_t(b^iB/n)}{nG(u)}.$$

By the monotonicity of $G(u)$, it follows that

$$\begin{aligned}
\left|\frac{\sum_{t=1}^{n} D_t(u)}{nG(u)}\right| &\le \left|\frac{\sum_{t=1}^{n} D_t(b^iB/n)}{nG(u)}\right| + \left|\frac{\sum_{t=1}^{n} \tilde{D}_t(b^iB/n,u)}{nG(u)}\right| \\
&\le \left|\frac{\sum_{t=1}^{n} D_t(b^iB/n)}{nG(b^iB/n)}\right| + \left|\frac{\sum_{t=1}^{n} \tilde{D}_t(b^iB/n,b^{i+1}B/n)}{nG(b^iB/n)}\right| \\
&\le \left|\sum_{t=1}^{n}\frac{D_t(b^iB/n)}{nG(b^iB/n)}\right| + \left|\frac{\mathbb{E}\big(\tilde{D}_t(b^iB/n,b^{i+1}B/n)\big)}{G(b^iB/n)}\right| \\
&\quad + \left|\sum_{t=1}^{n}\frac{\tilde{D}_t(b^iB/n,b^{i+1}B/n)-\mathbb{E}\big(\tilde{D}_t(b^iB/n,b^{i+1}B/n)\big)}{nG(b^iB/n)}\right|.
\end{aligned}$$

Thus

$$\begin{aligned}
\sup_{i\ge 0}\sup_{u\in R_i}\left|\frac{\sum_{t=1}^{n} D_t(u)}{nG(u)}\right| \le& \sup_{i\ge 0}\left|\sum_{t=1}^{n}\frac{D_t(b^iB/n)}{G(b^iB/n)}\right| + \sup_{i\ge 0}\left|\frac{\mathbb{E}\big(\tilde{D}_t(b^iB/n,b^{i+1}B/n)\big)}{G(b^iB/n)}\right| \\
&+ \sup_{i\ge 0}\left|\sum_{t=1}^{n}\frac{\tilde{D}_t(b^iB/n,b^{i+1}B/n)-\mathbb{E}\big(\tilde{D}_t(b^iB/n,b^{i+1}B/n)\big)}{nG(b^iB/n)}\right| \\
\equiv& J_1+J_2+J_3.
\end{aligned}$$

For any $\eta>0$, by (S.5.7) and (S.5.8), we have

$$\mathbb{P}(J_1>\eta) < \sum_{i\ge 0}\frac{H}{nG(b^iB/n)\eta^2} \le \frac{H}{B\eta^2(1-b^{-1})}, \tag{S.5.9}$$

$$\mathbb{P}(J_3>\eta) < \sup_{i}\frac{H(b-1)}{b^iB\eta^2} \le \frac{H(b-1)}{B\eta^2}. \tag{S.5.10}$$

By (S.5.1), it follows that

$$J_2 \le H(b-1). \tag{S.5.11}$$

For any $\gamma > 0$ and $\epsilon > 0$, we can choose $\eta > 0$ and $b > 1$ sufficiently small such that $2\eta + H(b-1) < \gamma$, and then choose sufficiently large B, such that $\frac{H}{B\eta^2(1-b^{-1})} + \frac{H(b-1)}{B\eta^2} < \epsilon$. By (S.5.9)-(S.5.11), we have

$$\mathbb{P}\Big(\sup_{B/n<u<\delta}\Big|\frac{\sum_{t=1}^n D_t(u)}{nG(u)}\Big| > \gamma\Big) = \mathbb{P}\Big(\sup_{i\ge 0}\sup_{u\in R_i}\Big|\frac{\sum_{t=1}^n D_t(u)}{nG(u)}\Big| > \gamma\Big) \le \mathbb{P}(J_1 + J_2 + J_3 > \gamma) \le \epsilon.$$

The proof is complete.

(v). For $p = 1$ here, let $\tilde{y}_{t-1}^2 = y_{t-1}^2 - \mathbb{E}y_{t-1}^2$ and

$$\tilde{H}_t(u_1, u_2) = |\tilde{y}_{t-1}^2|\tilde{I}_{t-1}(u_1, u_2).$$

Then we have

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\tilde{H}_1(u_1,u_2), \tilde{H}_j(u_1,u_2)\right)\right| \\
=&\left|\mathbb{E}\left\{\tilde{y}_0^2\tilde{I}_0(u_1,u_2)\left[\mathbb{E}\left(\tilde{y}_{j-1}^2\tilde{I}_{j-1}(u_1,u_2)\middle|\mathbf{x}_0\right) - \mathbb{E}\left(\tilde{y}_0^2\tilde{I}_0(u_1,u_2)\right)\right]\right\}\right| \\
=&\mathbb{E}\left\{y_0^2\tilde{I}_0(u_1,u_2)\left|\int \mathbb{E}(\tilde{y}^2|z)\tilde{I}(u_1,u_2)\left[\mathbb{P}^{j-1}(\mathbf{x}_0, dx) - \Pi(dx)\right]\right|\right\} \\
\le& C\mathbb{E}\left\{\tilde{y}_0^2\tilde{I}_0(u_1,u_2)\|\mathbb{P}^{j-1}(\mathbf{x}_0,\cdot) - \Pi(\cdot)\|_{\mathrm{v}}\right\} \\
\le& C\rho^{j-1}\mathbb{E}\left\{\tilde{y}_0^2\tilde{I}_0(u_1,u_2)\left(1 + |y_0| + |z_0| + |w_0|\right)\right\} \\
=&O\left(\rho^j\left[G(u_2) - G(u_1)\right]\right).
\end{aligned}$$

We complete the proof by following similar arguments as in (iv). □

Lemma S.5.2. *If the conditions in Theorem 3.2 hold, then, for any* $0 < B < \infty$,

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\le B/n}\left\|\frac{\partial l_n(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}} - \frac{\partial l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\right\| = O_p(n^{-1}), \tag{S.5.12}$$

$$\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\le B/n}\left\|\frac{\partial^2 l_n(\boldsymbol{\beta},\boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} - \frac{\partial^2 l_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\| = O_p(n^{-1}). \tag{S.5.13}$$

Proof. By a direct calculation, it follows that

$$\begin{aligned}
\frac{\partial e_t(\boldsymbol{\beta},\boldsymbol{\tau})}{\partial\boldsymbol{\beta}} =& -\left(\mathbf{y}_{t-1}'I_{1t}(\boldsymbol{\tau}), \mathbf{y}_{t-1}'I_{2t}(\boldsymbol{\tau}), \mathbf{y}_{t-1}'I_{3t}(\boldsymbol{\tau}), \mathbf{y}_{t-1}'I_{4t}(\boldsymbol{\tau})\right)', \\
\frac{\partial^2 e_t^2(\boldsymbol{\beta},\boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} =& 2\begin{pmatrix} I_{1t}(\boldsymbol{\tau}) & & & \\ & I_{2t}(\boldsymbol{\tau}) & & \\ & & I_{3t}(\boldsymbol{\tau}) & \\ & & & I_{4t}(\boldsymbol{\tau})\end{pmatrix} \otimes (\mathbf{y}_{t-1}\mathbf{y}_{t-1}'),
\end{aligned}$$

where $\otimes$ is the Kronecker product. Note that

$$\begin{aligned}\Big\|\frac{\partial e_t^2(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}-\frac{\partial e_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\Big\| \leq & 2|e_t(\boldsymbol{\theta}_0)|\Big\|\frac{\partial e_t(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}-\frac{\partial e_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\Big\| \\ & + 2|e_t(\boldsymbol{\beta}_0,\boldsymbol{\tau})-e_t(\boldsymbol{\theta}_0)|\Big\|\frac{\partial e_t(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}\Big\|.\end{aligned} \tag{S.5.14}$$

Let $\Lambda_{1t} = I(r\wedge r_0 < z_{t-1} \leq r\vee r_0)$ and $\Lambda_{2t} = I(s\wedge s_0 < w_{t-1} \leq s\vee s_0)$. Recall $\Lambda_t = I(\{r\wedge r_0 < z_{t-1}\leq r\vee r_0\}\cup\{s\wedge s_0 < w_{t-1}\leq s\vee s_0\})$. Then we have

$$\begin{aligned}\Big\|\frac{\partial e_t(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}-\frac{\partial e_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\Big\| &\leq 2\|\mathbf{y}_{t-1}\|\Lambda_t, \\ |e_t(\boldsymbol{\beta}_0,\boldsymbol{\tau})-e_t(\boldsymbol{\theta}_0)| &\leq C\|\mathbf{y}_{t-1}\|\Lambda_t.\end{aligned} \tag{S.5.15}$$

By Assumption 3.4, (S.5.14)-(S.5.15) and a conditional argument, it follows that

$$\begin{aligned}&\mathbb{E}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\leq\eta}\Big\|\frac{\partial e_t^2(\boldsymbol{\beta}_0,\boldsymbol{\tau})}{\partial\boldsymbol{\beta}}-\frac{\partial e_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}}\Big\| \\ &\leq \mathbb{E}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\leq\eta} C(|e_t(\boldsymbol{\theta}_0)|\|\mathbf{y}_{t-1}\|+\|\mathbf{y}_{t-1}\|^2)\Lambda_t \\ &\leq \mathbb{E}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\leq\eta} C(|e_t(\boldsymbol{\theta}_0)|\|\mathbf{y}_{t-1}\|+\|\mathbf{y}_{t-1}\|^2)(\Lambda_{1t}+\Lambda_{2t}) \\ &\leq C\mathbb{E}\sup_{\|\boldsymbol{\tau}-\boldsymbol{\tau}_0\|\leq\eta}\big\{\Lambda_{1t}\mathbb{E}[(\|\mathbf{y}_{t-1}\|+\|\mathbf{y}_{t-1}\|^2)\big|z_{t-1}]+\Lambda_{2t}\mathbb{E}[(\|\mathbf{y}_{t-1}\|+\|\mathbf{y}_{t-1}\|^2)\big|w_{t-1}]\big\} \\ &\leq C\eta.\end{aligned}$$

Let $\eta = B/n$, so (S.5.12) follows.
Using the similar technique, we have

$$\Big\|\frac{\partial^2 e_t^2(\boldsymbol{\beta},\boldsymbol{\tau})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}-\frac{\partial^2 e_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\Big\| \leq C\|\mathbf{y}_{t-1}\mathbf{y}_{t-1}'\|\Lambda_t,$$

so that (S.5.13) can be proved. □

References

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*, number 3 in *Institute of Mathematical Statistics Monographs*, Cambridge University Press, Cambridge.

Brown, B. M. (1971). Martingale central limit theorems, *The Annals of Mathematical Statistics* **42**(1): 59–66.

Cai, Z. (2002). Regression quantiles for time series, *Econometric Theory* **18**(1): 169–192.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *The Annals of Statistics* **21**(1): 520–533.

Chen, H., Chong, T. T. L. and Bai, J. (2012). Theory and applications of TAR model with two threshold variables, *Econometric Reviews* **31**(2): 142–170.

De Gooijer, J. G. (2001). Cross-validation criteria for SETAR model selection, *Journal of Time Series Analysis* **22**(3): 267–281.

Gonzalo, J. and Pitarakis, J. Y. (2002). Estimation and model selection based inference in single and multiple threshold models, *Journal of Econometrics* **110**(2): 319–352.

Hamaker, E. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models, *Journal of Mathematical Psychology* **53**(6): 518–529.

Hansen, B. E. (1997). Inference in TAR models, *Studies in Nonlinear Dynamics & Econometrics* **2**(1): 1–16.

Hansen, B. E. (2000). Sample splitting and threshold estimation, *Econometrica* **68**(3): 575–603.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Univ of California Press, Berkeley, pp. 221–233.

Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models, *Journal of Econometrics* **167**(1): 240–253.

Li, D., Ling, S. and Li, W. K. (2013). Asymptotic theory on the least squares estimation of threshold moving-average models, *Econometric Theory* **29**(3): 482–516.

Li, D., Ling, S., Tong, H. and Yang, G. (2019). On brownian motion approximation of compound poisson processes with applications to threshold models, *Advances in Decision Sciences* **23**(2): 164–191.

Peña, D. and Rodriguez, J. (2005). Detecting nonlinearity in time series by model selection criteria, *International Journal of Forecasting* **21**(4): 731–748.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Seijo, E. and Sen, B. (2011). A continuous mapping theorem for the smallest argmax functional, *Electronic Journal of Statistics* **5**: 421–439.

Wong, C. S. and Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models, *Journal of Time Series Analysis* **19**(1): 113–124.

Yu, P. and Phillips, P. C. (2018). Threshold regression asymptotics: From the compound Poisson process to two-sided Brownian motion, *Economics Letters* **172**: 123–126.