Modelling time series of counts with hysteresis

Xintong Ma[1], Dong Li[1] and Howell Tong[1,2,3]

[1] *Tsinghua University,* [2] *Xiamen University*

and [3] *The London School of Economics and Political Science*

Supplementary Material

This Supplementary Material contains simulation study, additional details for real data analysis results, and proofs of all theorems in the article with some useful technical lemmas.

S1 Simulation Studies

S1.1 Performance of estimators

To evaluate the finite-sample performance of $\widehat{\theta}_n$, we take two sets of parameters for Monte Carlo simulation of the two models.

For the BPART model, observations are generated from the model with true parameters $\theta_0^{\mathrm{b}} = (0.5, 0.6, 0.4, 0.2, 0.4, 0.5, 3, 6)^T$ and $\theta_0^{\mathrm{b}} = (0.6, 0.8, 0.7, 0.4, 0.2, 0.2, 4, 7)^T$, respectively.

For the HPART model, observations are generated from the model with true parameters $\theta_0^{\mathrm{h}} = (0.5, 0.6, 0.4, 0.2, 0.4, 0.5, 3, 6, 0)^T$ and $\theta_0^{\mathrm{h}} = (0.6, 0.8, 0.7,$

$0.4, 0.2, 0.2, 4, 7, -1)^T$, respectively. We adopt the same parameters ϑ_0 in the two models to better learn about the performance in parameter estimation.

We set $\gamma_1 = 0.2$, $\gamma_2 = 0.8$ and the threshold bounds are determined by the γ_i-quantiles of $\{y_t\}_{t=1}^n$ and $\{\Delta y_t\}_{t=2}^n$. The length of observations is $n = 500$, 1000, and 2000, respectively. The tables below summarize the empirical mean (EM), the empirical variance $\mathrm{cov}(\widehat{\vartheta}_n)$ (EV), the sample mean of $\widehat{G}^{-1}/n$ (SG) and the ratios EV/EM (V/M) based on 1000 replications.

Table S.1: Simulation 1 for the BPART model

n	Description	ω_1	α_1	β_1	ω_2	α_2	β_2	r	s
	θ_0^{b}	0.5	0.6	0.4	0.2	0.4	0.5	3	6
500	EM	0.593	0.598	0.382	0.447	0.394	0.462	3.097	5.956
	EV	0.105	0.007	0.009	0.376	0.007	0.015	0.302	0.272
	SG	0.082	0.005	0.008	0.393	0.006	0.014	/	/
	V/M	0.176	0.011	0.022	0.840	0.017	0.031	0.097	0.046
1000	EM	0.546	0.599	0.391	0.325	0.395	0.484	3.043	6.001
	EV	0.038	0.002	0.004	0.129	0.003	0.006	0.087	0.023
	SG	0.035	0.002	0.004	0.158	0.003	0.006	/	/
	V/M	0.070	0.004	0.009	0.398	0.006	0.012	0.029	0.004
2000	EM	0.518	0.602	0.395	0.265	0.395	0.494	3	5.999
	EV	0.016	0.001	0.002	0.047	0.001	0.002	0.002	0.001
	SG	0.016	0.001	0.002	0.070	0.001	0.003	/	/
	V/M	0.031	0.002	0.004	0.177	0.003	0.005	0.001	0.000

From Tables S.1-S.4, we can see that $\widehat{\tau}_n$ converges to τ_0 generally. It

Table S.2: Simulation 2 for the BPART model

n	Description	ω_1	α_1	β_1	ω_2	α_2	β_2	r	s
	θ_0^{b}	0.6	0.8	0.7	0.4	0.2	0.2	4	7
500	EM	0.569	0.785	0.719	0.461	0.192	0.201	4	7
	EV	0.086	0.008	0.008	0.120	0.004	0.005	0	0
	SG	0.092	0.007	0.008	0.165	0.004	0.004	/	/
	V/M	0.150	0.010	0.012	0.261	0.021	0.023	0	0
1000	EM	0.572	0.791	0.713	0.423	0.198	0.200	4	7
	EV	0.044	0.004	0.004	0.074	0.002	0.002	0	0
	SG	0.045	0.004	0.004	0.082	0.002	0.002	/	/
	V/M	0.077	0.005	0.005	0.175	0.010	0.011	0	0
2000	EM	0.598	0.798	0.702	0.410	0.198	0.202	4	7
	EV	0.022	0.002	0.002	0.038	0.001	0.001	0	0
	SG	0.022	0.002	0.002	0.041	0.001	0.001	/	/
	V/M	0.036	0.002	0.003	0.093	0.005	0.005	0	0

seems that the speed of convergence heavily depends on other parameters. Specifically, corresponding to different sets of true parameters, Tables S.1 and S.3 show that even when n is as large as 2000, $\widehat{r}_n$ fails to hit r_0 exactly, while Tables S.2 and S.4 show that $(\widehat{r}_n, \widehat{s}_n)$ hits (r_0, s_0) exactly when $n = 500$. This phenomenon is also observed in the SETPAR model; see Wang et al. (2014). Overall, the asymptotic results of the estimates of the parameters are confirmed in both examples. Specifically, the means of the estimates of the parameters are close to the true values of the parameters and the accuracy increases as the length of observations increases.

Table S.3: Simulation 1 for the HPART model

n	Description	ω_1	α_1	β_1	ω_2	α_2	β_2	r	s	c
	θ_0^{h}	0.5	0.6	0.4	0.2	0.4	0.5	3	6	0
500	EM	0.595	0.593	0.384	0.498	0.398	0.448	3.333	5.956	-0.023
	EV	0.107	0.009	0.009	0.586	0.009	0.018	0.707	0.416	0.947
	SG	0.087	0.005	0.008	0.593	0.006	0.016	/	/	/
	V/M	0.180	0.015	0.023	1.175	0.022	0.040	0.212	0.070	-41.19
1000	EM	0.558	0.598	0.388	0.331	0.399	0.476	3.212	6.008	-0.068
	EV	0.041	0.003	0.004	0.203	0.003	0.007	0.431	0.078	0.436
	SG	0.038	0.002	0.004	0.243	0.003	0.007	/	/	/
	V/M	0.074	0.005	0.011	0.613	0.009	0.014	0.134	0.013	-6.409
2000	EM	0.535	0.597	0.394	0.280	0.398	0.488	3.094	6	-0.028
	EV	0.019	0.001	0.002	0.080	0.001	0.003	0.151	0.004	0.105
	SG	0.018	0.001	0.002	0.103	0.001	0.003	/	/	/
	V/M	0.036	0.002	0.005	0.287	0.003	0.006	0.049	0.001	-3.761

Table S.4: Simulation 2 for the HPART model

n	Description	ω_1	α_1	β_1	ω_2	α_2	β_2	r	s	c
	θ_0^{h}	0.6	0.8	0.7	0.4	0.2	0.2	4	7	-1
500	EM	0.598	0.792	0.709	0.413	0.198	0.199	4	7	-1
	EV	0.090	0.008	0.008	0.119	0.003	0.004	0	0	0
	SG	0.099	0.007	0.008	0.182	0.004	0.004	/	/	/
	V/M	0.151	0.009	0.011	0.287	0.018	0.021	0	0	0
1000	EM	0.598	0.795	0.706	0.406	0.197	0.202	4	7	-1
	EV	0.048	0.004	0.004	0.080	0.002	0.002	0	0	0
	SG	0.049	0.003	0.004	0.091	0.002	0.002	/	/	/
	V/M	0.080	0.004	0.006	0.196	0.009	0.010	0	0	0
2000	EM	0.596	0.796	0.704	0.399	0.198	0.202	4	7	-1
	EV	0.024	0.002	0.002	0.045	0.001	0.001	0	0	0
	SG	0.024	0.002	0.002	0.045	0.001	0.001	/	/	/
	V/M	0.041	0.002	0.003	0.112	0.005	0.005	0	0	0

However, for the intercept parameter ω_2 estimate, there is apparently a large variance-to-mean ratio in both examples, compared with the other parameter estimates. This phenomenon also exists in the PAR model, the SETPAR model, and even in the classical GARCH(1,1) model. So far, no explanation has been given in the literature.

S1.2 Performance of tests

We evaluate the performance of the above tests through Monte Carlo experiments. Observations are generated with the same true parameters as above except that we set $c_0 = 0, -1, 1$ for the HPART model. Here, $n =$ 500, 1000, 2000 and 4000, respectively. We set the significance levels at 0.1, 0.05, and 0.01. The corresponding critical values for testing H_0 are simulated with 20,000 replications, which are $(2.5582, 3.3017, 5.2140)^T$ and $(4.2236, 5.2382, 7.6855)^T$ respectively. Tables below summarize the size and power based on 1000 replications.

When testing H_0, the tables below show that the size and the power of the test approach their desired values under different scenarios, as n increases. For example, Table S.6 shows satisfactory results in terms of power. However, the situation with Table S.5 is different. Although the size appears to be fairly well matched, the power behaviour is much less

Table S.5: Simulation 1 for testing H_0

	Data	c_0	α	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
size	BPART		0.10	0.164	0.124	0.114	0.101
			0.05	0.095	0.069	0.057	0.050
			0.01	0.028	0.018	0.012	0.008
power	HPART	0	0.10	0.323	0.451	0.707	0.899
			0.05	0.212	0.336	0.600	0.827
			0.01	0.068	0.152	0.391	0.664
power	HPART	-1	0.10	0.341	0.551	0.829	0.992
			0.05	0.221	0.412	0.741	0.975
			0.01	0.075	0.168	0.514	0.875
power	HPART	1	0.10	0.378	0.594	0.792	0.896
			0.05	0.272	0.473	0.724	0.856
			0.01	0.099	0.262	0.562	0.778

Table S.6: Simulation 2 for testing H_0

	Data	c_0	α	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
size	BPART		0.10	0.129	0.130	0.121	0.115
			0.05	0.081	0.079	0.064	0.062
			0.01	0.035	0.021	0.023	0.013
power	HPART	0	0.10	0.992	1	1	1
			0.05	0.991	1	1	1
			0.01	0.980	1	1	1
power	HPART	-1	0.10	0.988	1	1	1
			0.05	0.985	1	1	1
			0.01	0.968	1	1	1
power	HPART	1	0.10	0.991	1	1	1
			0.05	0.988	1	1	1
			0.01	0.969	1	1	1

satisfactory before the sample reaches 4000. Further, values of c_0 appear to exert some effect on the power.

The results in Table S.7 and Table S.8 for testing $\widetilde{H}_0$ are generally better than those for testing H_0 discussed previously.

Table S.7: Simulation 1 for testing $\widetilde{H}_0$

	Data	c_0	α	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
size	HPART	0	0.10	0.088	0.098	0.092	0.113
			0.05	0.044	0.040	0.045	0.051
			0.01	0.010	0.006	0.008	0.012
size	HPART	-1	0.10	0.079	0.096	0.091	0.101
			0.05	0.037	0.048	0.044	0.047
			0.01	0.007	0.006	0.005	0.011
size	HPART	1	0.10	0.077	0.099	0.093	0.109
			0.05	0.045	0.042	0.051	0.057
			0.01	0.014	0.010	0.006	0.012
power	BPART		0.10	0.320	0.632	0.837	0.967
			0.05	0.242	0.555	0.803	0.961
			0.01	0.121	0.431	0.725	0.933

Table S.8: Simulation 2 for testing $\widetilde{H}_0$

	Data	c_0	α	$n = 500$	$n = 1000$	$n = 2000$	$n = 4000$
size	HPART	0	0.10	0.103	0.112	0.114	0.090
			0.05	0.052	0.054	0.057	0.045
			0.01	0.011	0.013	0.010	0.006
size	HPART	-1	0.10	0.102	0.111	0.119	0.098
			0.05	0.053	0.044	0.050	0.044
			0.01	0.011	0.008	0.007	0.010
size	HPART	1	0.10	0.108	0.107	0.096	0.108
			0.05	0.064	0.060	0.043	0.061
			0.01	0.019	0.018	0.015	0.017
power	BPART		0.10	0.990	1	1	1
			0.05	0.986	1	1	1
			0.01	0.980	1	1	1

S2 Additional details for real data analysis

To verify the model adequacy, we conduct model diagnostic checking in this section. We also provide additional ID cards analysis for the SETPAR model.

S2.1 Diagnostic checking and ID cards analysis in Subsection 5.1

Table S.9 reports the values of AIC for model fittings. From the table, the BPART and HPART models outperform the standard PAR and SETPAR ones. Fig. S.2 displays the ACF plot of the Pearson residuals and Table S.10 summarizes their descriptive statistics. Combined with p-values of the Ljung–Box test plotted in Fig. S.1, the Pearson residual series for all four models show no significant serial dependence, indicating the adequacy of all model fittings.

Table S.9: Summary of model adequacy

	PAR	SETPAR	BPART	HPART
AIC	980.96	988.03	980.07	978.82

Table S.10: Statistics summary of the Pearson residuals

	PAR	SETPAR	BPART	HPART
Mean	0.01	0.01	0.01	0.01
Standard error	1.25	1.25	1.22	1.22
Skewness	0.41	0.43	0.44	0.47
Excess kurtosis	0.26	0.40	0.21	0.20

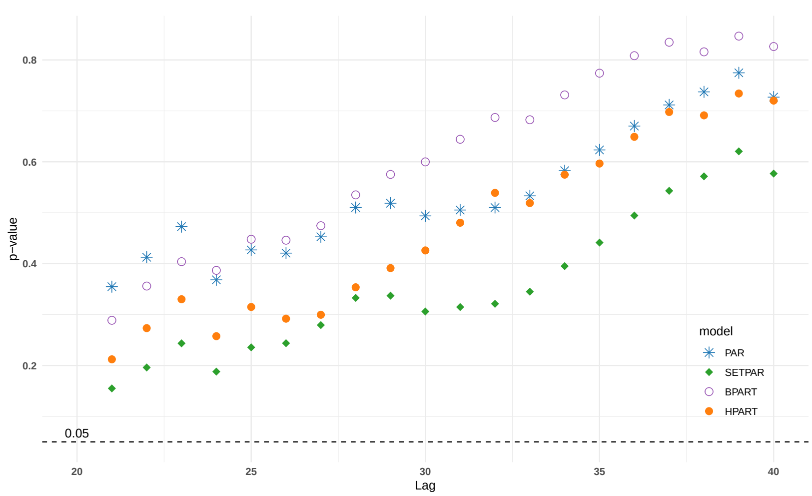
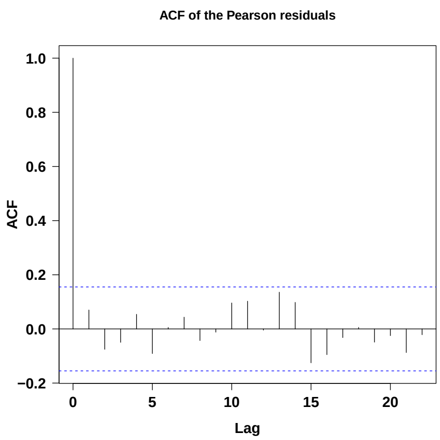
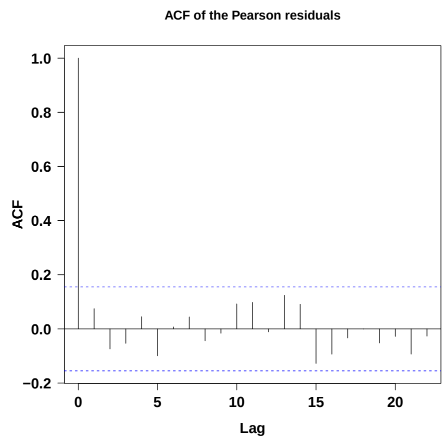


Figure S.1: P-values of Ljung-Box Test.

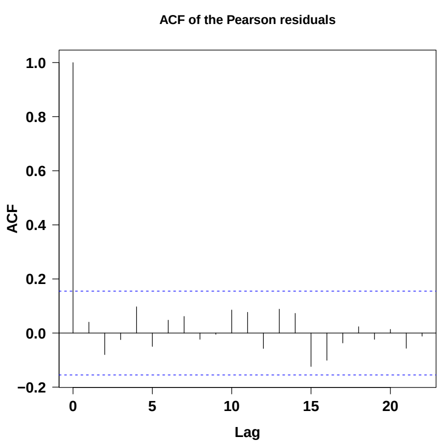
With respect to the ID card analysis, Fig. S.3 shows the binary sequence



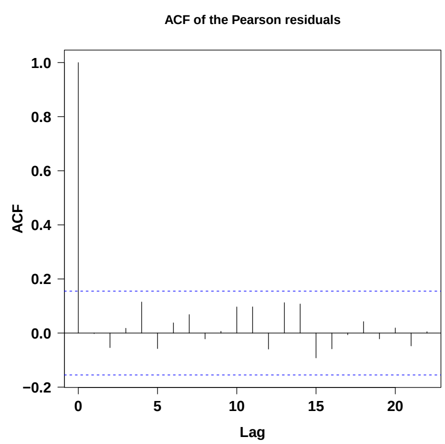
(a) PAR



(b) SETPAR



(c) BPART



(d) HPART

Figure S.2: ACF of the Pearson residuals of the escape custody data fitted by four different models of datum's ID cards generated by the three models separately. Based on the BIC, the ID sequences for the SETPAR model follow a zero-order Markov chain, while the BPART and HPART models display first-order Markov behavior, thereby distinguishing the SETPAR model from the other two.

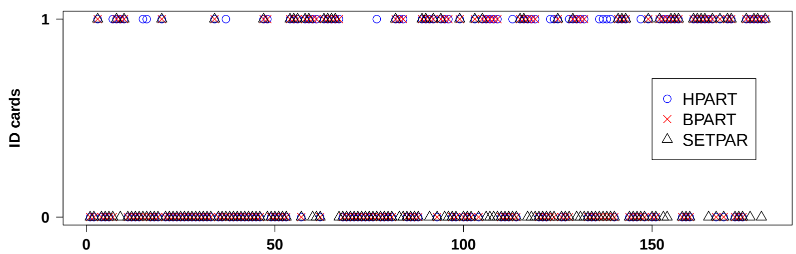


Figure S.3: The sequence of datum's ID cards

S2.2 Diagnostic checking and ID cards analysis in Subsection 5.2

For the Hepatitis B data, Tables S.11- S.12 and Fig. S.4-S.5 show that we reach similar diagnostic conclusions to those for the escape custody data.

Table S.11: Summary of model adequacy

	PAR	SETPAR	BPART	HPART
AIC	511.39	521.08	509.22	510.21

Fig. S.6 illustrates the binary sequence of datum's ID cards generated by the three models separately. Based on the BIC, the ID sequences for all three models follow a Markov chain order of 0, implying that they can

Table S.12: Statistics summary of the Pearson residuals

	PAR	SETPAR	BPART	HPART
Mean	0.07	-0.01	-0.01	0.00
Standard error	1.32	1.32	1.26	1.26
Skewness	0.56	0.58	0.61	0.60
Excess kurtosis	-0.26	-0.23	-0.11	-0.01

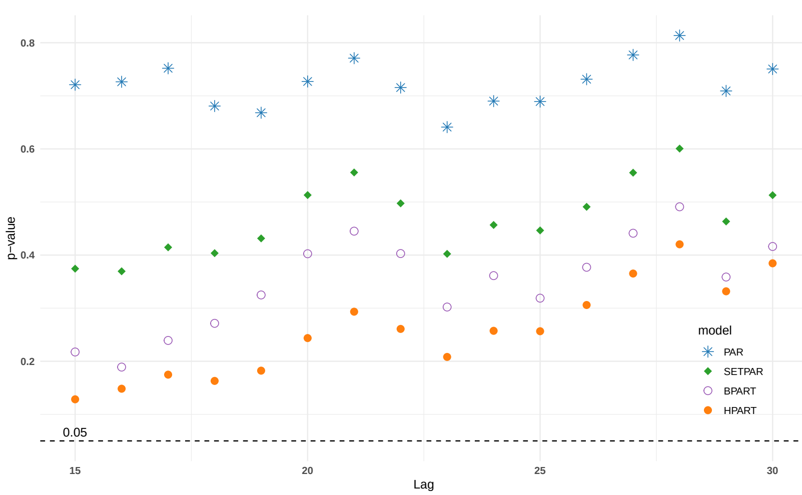
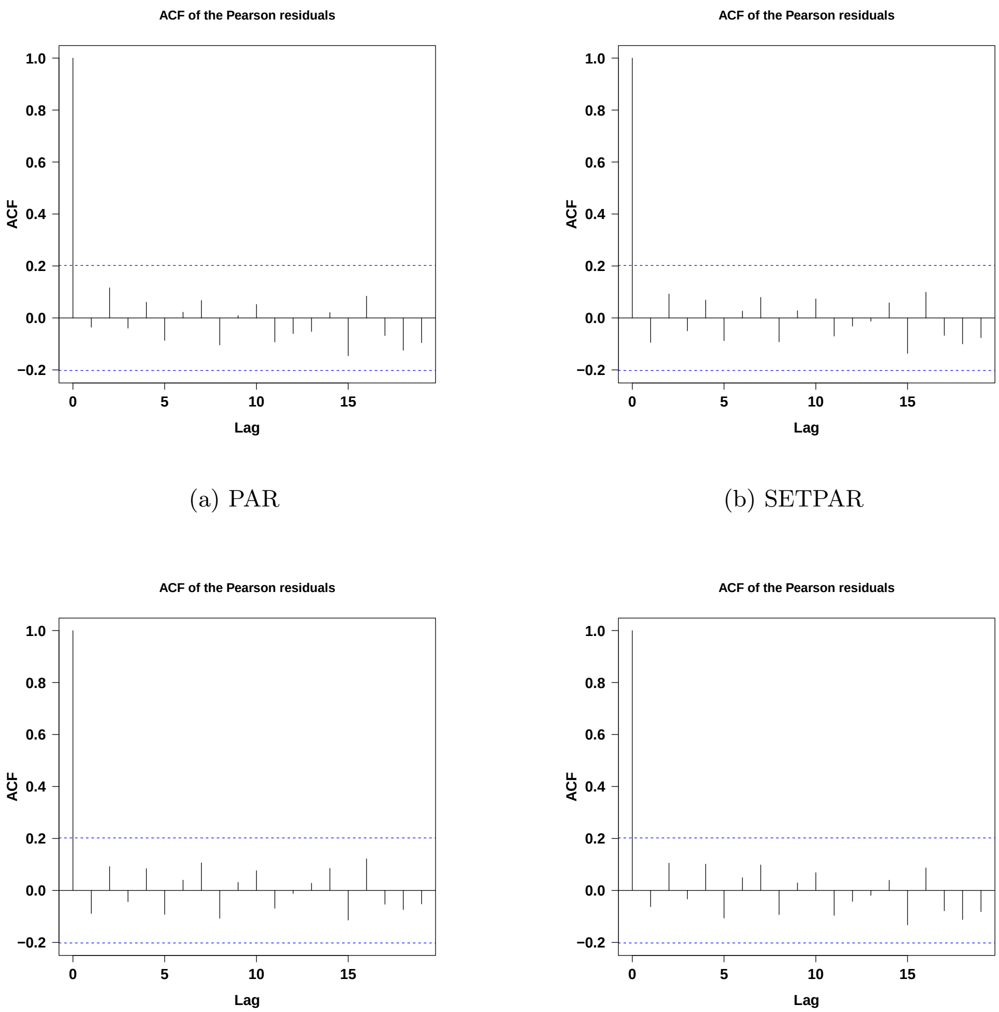


Figure S.4: P-values of Ljung-Box Test.



(a) PAR

(b) SETPAR

(c) BPART

(d) HPART

Figure S.5: ACF of the Pearson residuals of the Hepatitis B data fitted by four different models be treated as independent data. Table S.13 and S.14 present the contingency table of ID cards derived from the SETPAR and the BPART/HPART models. The p-values of the exact binomial test are 3×10^{-8} and 2×10^{-6}, respectively, indicating the SETPAR model differs significantly from the other two models.

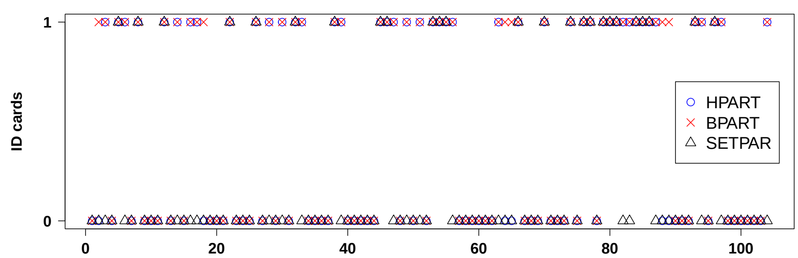


Figure S.6: The sequence of datum's ID cards

Table S.13: Contingency Table of ID cards

	BPART:ID=1	BPART:ID=0	total
SETPAR:ID=1	25	0	25
SETPAR:ID=0	26	53	79
total	51	53	104

Table S.14: Contingency Table of ID cards

	HPART:ID=1	HPART:ID=0	total
SETPAR:ID=1	25	0	25
SETPAR:ID=0	20	59	79
total	45	59	104

S3 Proofs of Theorems

Since the switching mechanism for data within the buffer regime of the BPART model depends on an infinite past which implies a more complex case. Care is necessary to demonstrate the strong consistency and asymptotic normality of the MLE of the model parameters. We give a full proof for the BPART model, noting that the proof for the HPART model is similar.

S3.1 Proof of Theorem 1

Consider an arbitrary (small) open neighborhood of $\theta_0 \in \Theta$, say V, then for any $\theta' \in V^c \cap \Theta$, we have $E(\ell_t(\theta')) < E(\ell_t(\theta_0))$ by Lemma 8. Lemma 7 implies that $E(\ell_t(\theta))$ is continuous in θ. Denote

$$\kappa = E(\ell_t(\theta_0)) - \sup_{\theta' \in V^c \cap \Theta} E(\ell_t(\theta')) > 0.$$

For any $\theta \in V^c \cap \Theta$, by Lemma 7 again, there exists an $\eta_\theta > 0$ such that

$$E\Big(\sup_{\theta'' \in V_{\eta_\theta}(\theta)} \ell_t(\theta'')\Big) < E(\ell_t(\theta)) + \kappa/6.$$

Since $V^c \cap \Theta$ is compact, there exists a finite covering $\{V_{\eta_{\theta_j}}(\theta_j),\ \theta_j \in V^c \cap \Theta, j = 1, ..., T\}$ of $V^c \cap \Theta$ such that $V^c \cap \Theta \subset \bigcup_{j=1}^{T} V_{\eta_{\theta_j}}(\theta_j)$. By Lemma 6, we have a.s., $1 \le j \le T$,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \frac{1}{n}\widetilde{L}_n(\theta^*) \\
&\le \overline{\lim_{n\to\infty}} \sup_{\vartheta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \frac{1}{n}L_n(\theta^*) + \overline{\lim_{n\to\infty}} \sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \frac{1}{n}|\widetilde{L}_n(\theta^*) - L_n(\theta^*)| \\
&\le \overline{\lim_{n\to\infty}} \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \ell_t(\theta^*).
\end{aligned}$$

Thus, we have a.s. for sufficiently large n,

$$\begin{aligned}
\sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\theta^*) &\le \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \ell_t(\theta^*) + \kappa/6 \\
&\le E\Big(\sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)} \ell_t(\theta^*)\Big) + \kappa/3 \\
&\le E(\ell_t(\theta_j)) + \kappa/2 \le E(\ell_t(\theta_0)) - \kappa/2,
\end{aligned}$$

and

$$\sup_{\theta\in V} \frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\theta) \ge \frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\theta_0) \ge \frac{1}{n}\sum_{t=1}^{n}\ell_t(\theta_0) - \kappa/6 \ge E(\ell_t(\theta_0)) - \kappa/3.$$

Thus, for any neighborhood V of θ_0, it follows that for sufficiently large n,

$$\sup_{\theta^* \in V_{\eta_{\theta_j}}(\theta_j)\cap\Theta} \frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\theta^*) \le \sup_{\theta\in V}\frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\theta),$$

which implies $\widehat{\theta}_n \in V$ a.s. By the arbitrariness of V, we have $\widehat{\theta}_n \to \theta_0$ a.s. The proof is complete. ■

S3.2 Proof of Theorem 2

By the Taylor expansion, it follows that

$$0 = n^{-1/2}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\widehat{\vartheta}_n)}{\partial\vartheta} = n^{-1/2}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta} + \Big\{n^{-1}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\vartheta_*)}{\partial\vartheta\partial\vartheta^T}\Big\}\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0),$$

where ϑ_* satisfies that $\|\vartheta_* - \vartheta_0\| \le \|\widehat{\vartheta}_n - \vartheta_0\|$. It suffices to prove that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta} \stackrel{d}{\longrightarrow} N(0,G) \quad \text{and} \quad n^{-1}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\vartheta_*)}{\partial\vartheta\partial\vartheta^T} \stackrel{p}{\longrightarrow} -G,$$

which can be proved by the following statements

(S1). $n^{-1/2}\sum_{t=1}^{n}\frac{\partial\ell_t(\vartheta_0)}{\partial\vartheta} \stackrel{d}{\longrightarrow} N(0,G).$

(S2). $\Big\|n^{-1/2}\sum_{t=1}^{n}\frac{\partial\ell_t(\vartheta_0)}{\partial\vartheta} - n^{-1/2}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta}\Big\| \stackrel{p}{\longrightarrow} 0,$

(S3). $E\Big\{\sup_{\vartheta\in\Theta_\vartheta}\frac{\partial^3\ell_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j\partial\vartheta^k}\Big\} < \infty, \quad i,j,k = 1,...,6.$

(S4). $\sup_{\vartheta\in\Theta_\vartheta}\Big\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\ell_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\vartheta)}{\partial\vartheta\partial\vartheta^T}\Big\| \stackrel{p}{\longrightarrow} 0,$

(S5). $\sup_{\|\vartheta-\vartheta_0\|<\epsilon}\Big\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\ell_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\ell_t(\vartheta_0)}{\partial\vartheta\partial\vartheta^T}\Big\| = O_p(\epsilon), \quad \forall\epsilon > 0.$

(S6). $n^{-1}\sum_{t=1}^{n}\frac{\partial^2\ell_t(\vartheta_0)}{\partial\vartheta\partial\vartheta^T} \stackrel{p}{\longrightarrow} -G,$

where ϑ^i, $i = 1,...,6$, is the ith component of ϑ, and '$\stackrel{p}{\longrightarrow}$' stands for convergence in probability. Next, we show (S1)-(S6), respectively.

First, note that $\{\partial\ell_t(\vartheta_0)/\partial\vartheta\}$ is a sequence of stationary and ergodic martingale differences since

$$E\Big\{\frac{\partial \ell_t(\vartheta_0)}{\partial\vartheta}\Big\} = E\Big\{\Big(\frac{y_t}{\lambda_t}-1\Big)\frac{\partial\lambda_t}{\partial\vartheta}\Big\} = E\Big[\frac{\partial\lambda_t}{\partial\vartheta}E\Big\{\Big(\frac{y_t}{\lambda_t}-1\Big)\ \Big|\ \mathcal{F}_{t-1}\Big\}\Big] = 0,$$

$$\mathrm{Var}\Big\{\frac{\partial \ell_t(\vartheta_0)}{\partial\vartheta}\Big\} = E\Big(\frac{1}{\lambda_t}\frac{\partial\lambda_t}{\partial\vartheta}\frac{\partial\lambda_t}{\partial\vartheta^T}\Big) = G.$$

By the Martingale Central Limit Theorem, (S1) holds.

Second, for (S2), using the explicit expressions of $\partial\lambda_t(\vartheta)/\partial\vartheta$ and $\partial\widetilde{\lambda}_t(\vartheta)/\partial\vartheta$, a simple calculation yields that $E(\sup_{\vartheta\in\Theta_\vartheta}\|\partial\lambda_t(\vartheta)/\partial\vartheta\|)<\infty$,

$$E\Big(\sup_{\vartheta\in\Theta_\vartheta}\Big\|\frac{\partial\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta}-\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta}\Big\|\Big)\le C\rho^t,\quad E\Big(\sup_{\vartheta\in\Theta_\vartheta}\Big\|\frac{\partial\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta}-\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta}\Big\|^2\Big)\le C\rho^t$$

where $\rho\in(0,1)$ and C is a generic constant that can change across lines. Since $E\{\sup_{\theta\in\Theta}|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)|^2\}=O(\rho^t)$, we have $E\big|\lambda_t^{-1}-\widetilde{\lambda}_t^{-1}\big|^2\le C\rho^t$, where $\lambda_t=\lambda_t(\theta_0)$ and $\widetilde{\lambda}_t=\widetilde{\lambda}_t(\theta_0)$. Note that

$$\Big\|\frac{\partial \ell_t(\vartheta_0)}{\partial\vartheta}-\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta}\Big\| = \Big\|y_t\Big\{\Big(\frac{1}{\lambda_t}-\frac{1}{\widetilde{\lambda}_t}\Big)\frac{\partial\lambda_t}{\partial\vartheta}+\frac{1}{\widetilde{\lambda}_t}\Big(\frac{\partial\lambda_t}{\partial\vartheta}-\frac{\partial\widetilde{\lambda}_t}{\partial\vartheta}\Big)\Big\}-\Big(\frac{\partial\lambda_t}{\partial\vartheta}-\frac{\partial\widetilde{\lambda}_t}{\partial\vartheta}\Big)\Big\|.$$

Since $E(y_t^2\|\partial\widetilde{\lambda}_t/\partial\vartheta\|^2)<\infty$, by the Minkowski inequality, it follows that, for any $\epsilon>0$,

$$\begin{aligned}
&P\Big\{\Big\|n^{-1/2}\sum_{t=1}^n\Big(\frac{\partial \ell_t(\vartheta_0)}{\partial\vartheta}-\frac{\partial\widetilde{\ell}_t(\vartheta_0)}{\partial\vartheta}\Big)\Big\|\ge\epsilon\Big\}\\
&\le (n^{1/2}\epsilon)^{-1}\sum_{t=1}^n\Big\{E\Big(y_t|\lambda_t^{-1}-\widetilde{\lambda}_t^{-1}|\Big\|\frac{\partial\lambda_t}{\partial\vartheta}\Big\|\Big)+C\rho^t\Big\}\\
&\le (n^{1/2}\epsilon)^{-1}\sum_{t=1}^n C\rho^t\Big[1+\Big\{E\Big(y_t^2\Big\|\frac{\partial\lambda_t}{\partial\vartheta}\Big\|^2\Big)\Big\}^{1/2}\Big]\\
&\to 0\quad \text{as } n\to\infty.
\end{aligned}$$

Thus, (S2) holds.

Third, for (S3), note that

$$\frac{\partial^3 \ell_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j\partial\vartheta^k} = \Big(-\frac{y_t}{\lambda_t^2(\vartheta)}\Big)\Big(\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^k} + \frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^j\partial\vartheta^k}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i} + \frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^k\partial\vartheta^i}\frac{\partial\lambda_t(\vartheta)}{\partial\partial\vartheta^j}\Big)$$
$$+ 2\Big(\frac{y_t}{\lambda_t^3(\vartheta)}\Big)\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^j}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^k} + \Big(\frac{y_t}{\lambda_t(\vartheta)} - 1\Big)\frac{\partial^3\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j\partial\vartheta^k}.$$

Take $\partial^i\lambda_t(\vartheta)/\partial\beta_1^i$ as an example, for $i = 1, 2, 3$,

$$\frac{\partial^i\lambda_t(\vartheta)}{\partial\beta_1^i} \le \sum_{k=1}^{\infty}\frac{(k-1)\cdots(k-i)}{\beta_1^i}\Big(\prod_{j=1}^{k-1} b_{t-j}\Big)a_{t-k},$$

where $b_{t-1} = \beta_1 R_t + \beta_2(1 - R_t)$ and

$$a_t = (\omega_1 + \alpha_1 y_t)R_{t+1}(\theta) + (\omega_2 + \alpha_2 y_t)\{1 - R_{t+1}(\theta)\}$$
$$= [\omega_1 R_{t+1}(\theta) + \omega_2\{1 - R_{t+1}(\theta)\}] + [\alpha_1 R_{t+1}(\theta) + \alpha_2\{1 - R_{t+1}(\theta)\}]y_t.$$

Denote $\underline{\beta} = \min_{\vartheta\in\Theta_\vartheta}\{\beta_1, \beta_2\} > 0$. Then

$$\frac{\partial^i\lambda_t(\vartheta)}{\partial\beta_1^i} \le \sum_{k=1}^{\infty}\frac{(k-1)\cdots(k-i)}{\underline{\beta}^i}\rho^{k-1}a_{t-k}.$$

Since $E(\sup_{\vartheta\in\Theta_\vartheta} a_t) \le CE(1 + y_t) < \infty$, we thus have that

$$E\Big\{\sup_{\vartheta\in\Theta_\vartheta}\frac{\partial^i\lambda_t(\vartheta)}{\partial\beta_1^i}\Big\} < \infty \quad \text{and} \quad E\Big\{\sup_{\vartheta\in\Theta_\vartheta}\frac{\partial^i\ell_t(\vartheta)}{\partial\beta_1^i}\Big\} < \infty.$$

Similarly, we can prove other cases. Thus, (S3) holds.

Forth, for (S4), it follows that, similar to (S2),

$$E\Big(\sup_{\vartheta\in\Theta_\vartheta}\Big\|\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} - \frac{\partial^2\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta\partial\vartheta^T}\Big\|\Big) \le C\rho^t, \quad E\Big(\sup_{\vartheta\in\Theta_\vartheta}\Big\|\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} - \frac{\partial^2\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta\partial\vartheta^T}\Big\|^2\Big) \le C\rho^t.$$

Using the expression

$$\frac{\partial^2 \ell_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} = \Big\{\frac{y_t}{\lambda_t(\vartheta)} - 1\Big\}\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta\partial\vartheta^T} - \frac{y_t}{\lambda_t^2}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^T},$$

we have that

$$\begin{aligned}\frac{\partial^2 \ell_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} - \frac{\partial^2 \widetilde{\ell}_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} &= y_t\Big[\Big\{\frac{1}{\lambda_t(\vartheta)} - \frac{1}{\widetilde{\lambda}_t(\vartheta)}\Big\}\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} + \frac{1}{\widetilde{\lambda}_t(\vartheta)}\Big\{\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} - \frac{\partial^2\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j}\Big\}\\
&\quad - \Big\{\frac{1}{\lambda_t^2(\vartheta)} - \frac{1}{\widetilde{\lambda}_t^2(\vartheta)}\Big\}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^j}\\
&\quad - \frac{1}{\widetilde{\lambda}_t^2(\vartheta)}\Big[\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i}\Big\{\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^j} - \frac{\partial\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta^j}\Big\} + \frac{\partial\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta^j}\Big\{\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i} - \frac{\partial\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta^i}\Big\}\Big]\Big]\\
&\quad - \Big\{\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} - \frac{\partial^2\widetilde{\lambda}_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j}\Big\}.\end{aligned}$$

Since

$$\begin{aligned}&E\Big(\sup_{\vartheta\in\Theta_\vartheta}\Big|\frac{\partial^2 \ell_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} - \frac{\partial^2 \widetilde{\ell}_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j}\Big|\Big)\\
&\le CE\Big[\sup_{\vartheta\in\Theta_\vartheta}\Big\{y_t\Big|\frac{1}{\lambda_t(\vartheta)} - \frac{1}{\widetilde{\lambda}_t(\vartheta)}\Big|\Big|\frac{\partial^2\lambda_t(\vartheta)}{\partial\vartheta^i\partial\vartheta^j} + \frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i}\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^j}\Big|\Big\}\Big]\\
&\quad + C\rho^t E\Big[\sup_{\vartheta\in\Theta_\vartheta}\Big\{y_t^2\Big|\frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^i} + \frac{\partial\lambda_t(\vartheta)}{\partial\vartheta^j}\Big|^2\Big\}\Big]^{1/2} + C\rho^t\{E(y_t^2)^{1/2} + 1\},\end{aligned}$$

similar to (S2), it is not hard to show (S4) holds.

Finally, (S5) holds by (S3) and the Taylor expansion. (S6) holds by the Law of Large Numbers for stationary and ergodic sequences. The proof is complete. ■

S3.3 Proof of Theorem 3

By Assumptions 1-4, the Martingale Central Limit Theorem implies that

$$\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial \vartheta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_t(\theta_0)}{\partial \vartheta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{y_t}{\lambda_t(\theta_0)} - 1\Big)\frac{\partial \lambda_t(\theta_0)}{\partial \vartheta} \xrightarrow{d} N(0, G).$$

By the Strong Law of Large Numbers, it follows that

$$-\frac{1}{n}\frac{\partial^2 L_n(\theta_0)}{\partial \vartheta \partial \vartheta^T} \to G, \quad \text{a.s.}$$

Then, we have

$$\begin{aligned}
\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) &= \Big(-\frac{1}{n}\frac{\partial^2 L_n(\theta_0)}{\partial \vartheta \partial \vartheta^T}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n(\theta_0)}{\partial \vartheta} + o_p(1) \\
&= G^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{y_t}{\lambda_t(\theta_0)} - 1\Big)\frac{\partial \lambda_t(\theta_0)}{\partial \vartheta} + o_p(1) \xrightarrow{d} N(0, G).
\end{aligned}$$

Under H_0, by Lemmas 9-11 and the consistency of parameters,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\frac{\partial \widetilde{L}(0, \hat{\theta}^{\mathrm{b}})}{\partial \delta} &= \frac{1}{\sqrt{n}}\frac{\partial L(0, \hat{\theta}^{\mathrm{b}})}{\partial \delta} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\frac{\partial L\left(0, \theta_0\right)}{\partial \delta} + \frac{1}{n}\frac{\partial^2 L\left(0, \xi_1\right)}{\partial \delta \partial \vartheta^T}\sqrt{n}(\hat{\vartheta}^{\mathrm{b}} - \vartheta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\frac{\partial L\left(0, \theta_0\right)}{\partial \delta} + \left\{\frac{1}{n}\frac{\partial^2 L\left(0, \theta_0\right)}{\partial \delta \partial \vartheta^T} + (\xi_1 - \theta_0)^T\frac{1}{n}\frac{\partial^3 L\left(0, \xi_2\right)}{\partial \delta \partial \vartheta \partial \vartheta^T}\right\}\sqrt{n}(\hat{\vartheta}^{\mathrm{b}} - \vartheta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\frac{\partial L\left(0, \theta_0\right)}{\partial \delta} + \frac{1}{n}\frac{\partial^2 L\left(0, \theta_0\right)}{\partial \delta \partial \vartheta^T}\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{y_t}{\lambda_t^{\mathrm{b}}(\theta_0)} - 1\Big)\frac{\partial \lambda_t^{\mathrm{b}}\left(\theta_0\right)}{\partial \vartheta} \\
&\quad + (\xi_1 - \theta_0)^T\frac{1}{n}\frac{\partial^3 L\left(0, \xi_2\right)}{\partial \delta \partial \vartheta \partial \vartheta^T}\sqrt{n}(\hat{\vartheta}^{\mathrm{b}} - \vartheta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{y_t}{\lambda_t^{\mathrm{b}}(\theta_0)} - 1\Big)\Big[\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\} + E\Big\{\frac{1}{n}\frac{\partial^2 L\left(0, \theta_0\right)}{\partial \delta \partial \vartheta^T}\Big\}\Sigma_1^{-1}\frac{\partial \lambda_t^{\mathrm{b}}\left(\theta_0\right)}{\partial \vartheta}\Big] + o_p(1). \\
\frac{1}{n}\frac{\partial^2 \widetilde{L}(0, \hat{\theta}^{\mathrm{b}})}{\partial \delta^2} &= \frac{1}{n}\frac{\partial^2 L(0, \hat{\theta}^{\mathrm{b}})}{\partial \delta^2} + o_p(1) = \frac{1}{n}\frac{\partial^2 L(0, \theta_0)}{\partial \delta^2} + o_p(1).
\end{aligned}$$

Denote

$$\sigma_1 = E\Big\{ -\frac{1}{n}\frac{\partial^2 L\,(0,\theta_0)}{\partial \delta^2}\Big\} = E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}^2\Big],$$

$$\sigma_2 = \sigma_1 - E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{b}}\,(\theta_0)}{\partial \vartheta^T}\Big]\Sigma_1^{-1}E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{b}}\,(\theta_0)}{\partial \vartheta}\Big].$$

We have

$$\frac{1}{\sqrt{n}}\frac{\partial L\,(0,\theta_0)}{\partial \delta} \xrightarrow{d} N(0,\sigma_2),$$
$$-\frac{1}{n}\frac{\partial^2 L\,(0,\theta_0)}{\partial^2 \delta} \xrightarrow{p} \sigma_1.$$

Further denote

$$\sigma_1(c_i,c_j) = E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\vartheta_0,r_0,s_0,c_i) - \lambda_t^{\mathrm{b}}(\theta_0)\}\{\lambda_t^{\mathrm{h}}(\vartheta_0,r_0,s_0,c_j) - \lambda_t^{\mathrm{b}}(\theta_0)\}\Big],$$

$$\sigma_2(c_i,c_j) = \sigma_1(c_i,c_j) - E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\vartheta_0,r_0,s_0,c_i) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{b}}\,(\theta_0)}{\partial \vartheta^T}\Big]$$
$$\times\,\Sigma_1^{-1}E\Big[\frac{1}{\lambda_t^{\mathrm{b}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\vartheta_0,r_0,s_0,c_j) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{b}}\,(\theta_0)}{\partial \vartheta}\Big].$$

Denote $\Sigma = (\sigma_2(c_i,c_j))$. By the Cramer-Wold device and the Martingale Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\left(\frac{\partial \widetilde{L}\big(0,\vartheta_0,r_0,s_0,c_1\big)}{\partial \delta},...,\frac{\partial \widetilde{L}\big(0,\vartheta_0,r_0,s_0,c_k\big)}{\partial \delta}\right)^T \xrightarrow{d} (Z(c_1),...,Z(c_k))^T,$$

where $(Z(c_1),...,Z(c_k))^T \sim N(0,\Sigma)$. Combining with the Law of Large Numbers, the Slutsky Theorem, and the continuous mapping theorem, we have

$$S_n := \max_{i\in[k]} T_n^{\mathrm{b}}(c_i) \xrightarrow{d} \max_{i\in[k]} \frac{Z^2(c_i)}{\sigma_1(c_i,c_i)}. \qquad \blacksquare$$

S3.4 Proof of Theorem 4

Similar to the proof of Theorem 3, under $\widetilde{H}_0$, denote

$$\sigma_1' = E\Big\{ -\frac{1}{n}\frac{\partial^2 L(1,\theta_0)}{\partial \delta^2}\Big\} = E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}^2\Big],$$

$$\sigma_2' = \sigma_1' - E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{h}}(\theta_0)}{\partial \vartheta^T}\Big]\Sigma_2^{-1} E\Big[\frac{1}{\lambda_t^{\mathrm{h}}(\theta_0)}\{\lambda_t^{\mathrm{h}}(\theta_0) - \lambda_t^{\mathrm{b}}(\theta_0)\}\frac{\partial \lambda_t^{\mathrm{h}}(\theta_0)}{\partial \vartheta}\Big].$$

It is not hard to prove

$$\frac{\hat{\sigma}_1'}{\hat{\sigma}_2'} T_n^{\mathrm{h}} \xrightarrow{d} \chi_1^2.$$

The proof is complete. ■

S4 Technical Lemmas

Lemma 1. *For fixed integers* $0 \le \underline{r} < \bar{s} < \infty$, *there exists a constant* $\rho \in (0,1)$ *such that*

$$\sup_{t\in\mathbb{Z}} E[I(\underline{r} < y_t \le \bar{s}) \mid \mathcal{F}_{t-1}] \le \rho.$$

Proof. Since $\lambda_t \ge \min\{\omega_{10},\ \omega_{20}\}$, we clearly have

$$E[I(\underline{r} < y_t \le \bar{s}) \mid \mathcal{F}_{t-1}] = P(\underline{r} < y_t \le \bar{s} \mid \mathcal{F}_{t-1}) = \Big(\sum_{k=\underline{r}+1}^{\bar{s}} \frac{\lambda_t^k}{k!}\Big)\exp(-\lambda_t)$$
$$\le \sup_{x\ge \min\{\omega_{10},\ \omega_{20}\}} f(x),$$

where

$$f(x) = \Big(\sum_{k=\underline{r}+1}^{\bar{s}} \frac{x^k}{k!}\Big)\exp(-x), \quad x \in [\min\{\omega_{10},\ \omega_{20}\},\ \infty).$$

It is well known that if $W \sim \text{Poisson}(x)$, then $(W-x)/\sqrt{x} \xrightarrow{d} N(0,1)$ as $x \to \infty$. Thus, for fixed integers $\underline{r}$ and $\bar{s}$, it follows that, as $x \to \infty$,

$$f(x) = P(\underline{r}+1 \le W \le \bar{s}) = P\Big(\frac{\underline{r}+1-x}{\sqrt{x}} \le \frac{W-x}{\sqrt{x}} \le \frac{\bar{s}-x}{\sqrt{x}}\Big)$$
$$\to P(N(0,1) = -\infty) = 0,$$

which implies that there exists a fintie $x_0 \in [\min\{\omega_{10},\ \omega_{20}\},\ \infty)$ such that

$$0 \le f(x) \le 1/2 \quad \text{for } x \ge x_0.$$

Clearly, $f(x)$ is continuous on the interval $[\min\{\omega_{10},\ \omega_{20}\},\ x_0]$ with $0 \le f(x) < 1$. Then,

$$\tilde{\rho} := \max_{x\in[\min\{\omega_{10},\ \omega_{20}\},\ x_0]} f(x) < 1.$$

Thus, there exists a constant $\rho \in (0,1)$ such that

$$\begin{aligned}\sup_{t\in\mathbb{Z}} E[I(\underline{r} < y_t \le \bar{s}) \mid \mathcal{F}_{t-1}] &\le \sup_{x\ge\min\{\omega_{10},\ \omega_{20}\}} f(x) \\ &\le \max\Big\{\max_{x\in[\min\{\omega_{10},\ \omega_{20}\},\ x_0]} f(x),\ 1/2\Big\} \\ &\le \max\{\tilde{\rho}, 1/2\} \le \rho < 1.\end{aligned}$$

The proof is complete. □

Lemma 2. *If Assumptions 1-3 hold, then,*

$$E\Big(\sup_{\theta\in\Theta} |R_t(\theta) - \widetilde{R}_t(\theta)|\Big) = O(\rho^t), \quad t \to \infty.$$

PROOF. By the definition of $\widetilde{R}_t$ and R_t, it follows that for $t \geq 1$,

$$\widetilde{R}_t(\theta) = \sum_{k=1}^{t} \Big\{ \prod_{j=1}^{k-1} I(r < y_{t-j} \leq s) \Big\} I(y_{t-k} \leq r) + \widetilde{R}_0(\theta) \prod_{j=1}^{t} I(r < y_{t-j} \leq s),$$

$$R_t(\theta) = \sum_{k=1}^{\infty} \Big\{ \prod_{j=1}^{k-1} I(r < y_{t-j} \leq s) \Big\} I(y_{t-k} \leq r).$$

Then, we have

$$R_t(\theta) - \widetilde{R}_t(\theta) = \sum_{k=t+1}^{\infty} \Big\{ \prod_{j=1}^{k-1} I(r < y_{t-j} \leq s) \Big\} I(y_{t-k} \leq r) - \widetilde{R}_0(\theta) \prod_{j=1}^{t} I(r < y_{t-j} \leq s).$$

By Lemma 1, a conditional argument yields that

$$\begin{aligned} E\Big\{ \sup_{\theta \in \Theta} |R_t(\theta) - \widetilde{R}_t(\theta)| \Big\} &\leq \sum_{k=t+1}^{\infty} E\Big\{ \prod_{j=1}^{k-1} I(\underline{r} < y_{t-j} \leq \overline{s}) \Big\} + E\Big\{ \prod_{j=1}^{t} I(\underline{r} < y_{t-j} \leq \overline{s}) \Big\} \\ &\leq \sum_{k=t+1}^{\infty} \rho^{k-1} + \rho^t = \frac{\rho^t}{1-\rho} + \rho^t = O(\rho^t). \end{aligned}$$

The proof is complete. □

Lemma 3. *[Lemma A.1 in Wang et al. (2014)] For a Poisson process* $\{N(u), u \geq 0\}$ *with unit rate,*

(1) $N(u)/u \to 1$ *a.s. as* $u \to \infty$.

(2) $\{(N(u)/u)^s, u \geq 1\}$ *is uniformly integrable for any integer* $s \geq 1$.

Following Fokianos et al. (2009), y_t can be rephrased by $y_t = N_t(\lambda_t)$, where $\{N_t(\cdot), t \in \mathbb{Z}\}$ is a sequence of independent Poisson processes with unit intensity. Thus, for the BPART model, $\{(\lambda_t, R_t)^T\}$ is a Markov chain. Denote its transition probability kernel by $\mathbf{P}$. For any function $V : \mathbb{R}_+ \times$

$\{0,1\} \to \mathbb{R}_+$, let $\mathbf{P}V(x,R) = E\{V(\lambda_1, R_1) \mid \lambda_0 = x, R_0 = R\}$. Similar to Lemma A.2 in Wang et al. (2014), we have

Lemma 4. *For any $p \geq 1$, let $V(x,R) = x^p + R^p$, $x > 0$, then*

$$\lim_{x\to\infty} \frac{\mathbf{P}V(x,R)}{V(x,R)} = (\alpha_2 + \beta_2)^p.$$

Proof. Clearly, a simple calculation gives that

$$\begin{aligned}
\frac{\mathbf{P}V(x,R)}{V(x,R)} &= \frac{E\{V(\lambda_1, R_1) \mid \lambda_0 = x, R_0 = R)\}}{V(x,R)} \\
&= \frac{E\{V(\lambda_1, R_1) \mid \lambda_0 = x, R_0 = R)\}}{x^p} \frac{x^p}{x^p + R^p} \\
&= E\Big[\Big(\frac{\omega_1}{x} + \alpha_1 \frac{y_0}{x} + \beta_1\Big)^p I(y_0 \leq r) \\
&\quad + \Big\{\Big(\frac{\omega_1}{x} + \alpha_1 \frac{y_0}{x} + \beta_1\Big)^p I(R = 1) \\
&\quad + \Big(\frac{\omega_2}{x} + \alpha_2 \frac{y_0}{x} + \beta_2\Big)^p I(R = 0)\Big\} I(r < y_0 \leq s) \\
&\quad + \Big(\frac{\omega_2}{x} + \alpha_2 \frac{y_0}{x} + \beta_2\Big)^p I(y_0 > s) + \frac{R_1^p}{x^p}\Big] \frac{x^p}{x^p + R^p} \\
&:= E\{h(x)\} \frac{x^p}{x^p + R^p}.
\end{aligned}$$

Note that $y_0|_{\lambda_0 = x} \sim \text{Poisson}(x)$. If $x \to \infty$, then $y_0/x = N_0(x)/x \to 1$ a.s. by Lemma 3 and $I(y_0 \leq r) \to 0$, $I(r < y_0 \leq s) \to 0$. Thus, we get $h(x) \to (\alpha_2 + \beta_2)^p$ a.s. For $x \geq 1$, $0 \leq h(x) \leq C(p)[1 + (y_0/x)^p]$. Then, $\{h(x), x \geq 1\}$ is uniformly integrable because of the uniform integrability of $\{(y_0/x)^p, x \geq 1\}$. The result holds. $\square$

Remark 1. For the HPART model, clearly, $\{(\lambda_t, y_{t-1})^T\}$ is a Markov chain.

For any $p \geq 1$, let $V(x_1, x_0) = x_1^p + x_0^p : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$, and $\mathbf{P}V(x_1, x_0) = E\{V(\lambda_2, y_1) \mid \lambda_1 = x_1, y_0 = x_0\}$. Then Lemma 4 still holds.

Lemma 5. *If Assumptions 1-3 hold, then there exists a constant $\rho \in (0,1)$ such that as $t \to \infty$,*

(i). $E\Big\{\sup_{\theta\in\Theta} \widetilde{\lambda}_{t-1}^p(\theta)\Big\} = O(1)$ *for any $p \geq 1$;*

(ii). $E\big[\sup_{\theta\in\Theta} |\lambda_t(\theta) - \widetilde{\lambda}_t(\theta)|\big] = O(\rho^t)$;

(iii). $E\big[\sup_{\theta\in\Theta} \{\lambda_t(\theta) - \widetilde{\lambda}_t(\theta)\}^2\big] = O(\rho^t)$;

(iv). $E\big[\sup_{\theta\in\Theta} y_t |\lambda_t(\theta) - \widetilde{\lambda}_t(\theta)|\big] = O(\rho^t)$.

Proof. $E(\lambda_t^p) = O(1)$ directly results from Lemma 4 and thus $E(y_t^p) = O(1)$ since y_t^p is a p-th degree polynomial of λ_t, similar to the proof of Corollary 1 in Wang et al. (2014).

Denote $\widetilde{W}_{t-1}(\theta) = \omega_1 I(\widetilde{R}_t(\theta) = 1) + \omega_2 I(\widetilde{R}_t(\theta) = 0)$, $\widetilde{A}_{t-1}(\theta) = \alpha_1 I(\widetilde{R}_t(\theta) = 1) + \alpha_2 I(\widetilde{R}_t(\theta) = 0)$, $\widetilde{B}_{t-1}(\theta) = \beta_1 I(\widetilde{R}_t(\theta) = 1) + \beta_2 I(\widetilde{R}_t(\theta) = 0)$ and $\bar{\omega} := \sup_{\theta\in\Theta} \max\{\omega_1, \omega_2\}$, $\bar{\alpha} := \sup_{\theta\in\Theta} \max\{\alpha_1, \alpha_2\}$, $\bar{\beta} := \sup_{\theta\in\Theta} \max\{\beta_1, \beta_2\}$. We have

$$\begin{aligned}
\widetilde{\lambda}_t(\theta) &= \widetilde{W}_{t-1}(\theta) + \widetilde{A}_{t-1}(\theta) y_{t-1} + \widetilde{B}_{t-1}(\theta)\widetilde{\lambda}_{t-1}(\theta) \\
&= \sum_{k=1}^{t} \prod_{j=1}^{k-1} \widetilde{B}_{t-j}(\theta)(\widetilde{W}_{t-k}(\theta) + \widetilde{A}_{t-k}(\theta) y_{t-k}) + \prod_{j=1}^{t} \widetilde{B}_{t-j}(\theta)\widetilde{\lambda}_0(\theta).
\end{aligned}$$

Thus,

$$\{E(\sup_{\theta\in\Theta}\widetilde{\lambda}_t^p(\theta))\}^{1/p} \le \sum_{k=1}^{t}\bar{\beta}^{k-1}(\bar{\omega}+\bar{\alpha}\{E(y_1^p)\}^{1/p}) + \bar{\beta}^t\{E(\widetilde{\lambda}_0(\theta)^p)\}^{1/p} < \infty.$$

From model (1.1), we have

$$\begin{aligned}
|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)| &= |(\omega_1+\alpha_1 y_{t-1})(R_t(\theta)-\widetilde{R}_t(\theta)) - (\omega_2+\alpha_2 y_{t-1})(R_t(\theta)-\widetilde{R}_t(\theta)) \\
&\quad + \{\beta_1 R_t(\theta)+\beta_2(1-R_t(\theta))\}\lambda_{t-1}(\theta) - \{\beta_1\widetilde{R}_t(\theta)+\beta_2(1-\widetilde{R}_t(\theta))\}\widetilde{\lambda}_{t-1}(\theta)| \\
&\le |(\omega_1+\alpha_1 y_{t-1})-(\omega_2+\alpha_2 y_{t-1})||R_t(\theta)-\widetilde{R}_t(\theta)| \\
&\quad + \rho|\lambda_{t-1}(\theta)-\widetilde{\lambda}_{t-1}(\theta)| + \rho\widetilde{\lambda}_{t-1}(\theta)|R_t(\theta)-\widetilde{R}_t(\theta)| \\
&= \rho|\lambda_{t-1}(\theta)-\widetilde{\lambda}_{t-1}(\theta)| + \epsilon_t(\theta) \\
&\le \sum_{j=0}^{t-1}\rho^j\epsilon_{t-j}(\theta) + \rho^t|\lambda_0(\theta)-\widetilde{\lambda}_0(\theta)|
\end{aligned}$$

where $\epsilon_t(\theta) = |(\omega_1+\alpha_1 y_{t-1})-(\omega_2+\alpha_2 y_{t-1})||R_t(\theta)-\widetilde{R}_t(\theta)|+\rho\widetilde{\lambda}_{t-1}(\theta)|R_t(\theta)-\widetilde{R}_t(\theta)|$.

By the Cauchy inequality and Lemma 2, it follows that

$$\begin{aligned}
& E\Big\{\sup_{\theta\in\Theta}|(\omega_1+\alpha_1 y_{t-1})-(\omega_2+\alpha_2 y_{t-1})||R_t(\theta)-\widetilde{R}_t(\theta)|\Big\} \\
&\le \{E|(\omega_1+\alpha_1 y_{t-1})-(\omega_2+\alpha_2 y_{t-1})|^2\}^{1/2}\Big\{E\Big(\sup_{\theta\in\Theta}|R_t(\theta)-\widetilde{R}_t(\theta)|^2\Big)\Big\}^{1/2} \\
&\le \{E(C+Cy_{t-1}+Cy_{t-1}^2)\}^{1/2}\Big\{E\Big(\sup_{\theta\in\Theta}|R_t(\theta)-\widetilde{R}_t(\theta)|\Big)\Big\}^{1/2} \\
&= O(\rho^t).
\end{aligned}$$

Similarly $E[\sup_{\theta\in\Theta}\{\widetilde{\lambda}_{t-1}(\theta)|R_t(\theta)-\widetilde{R}_t(\theta)|\}]=O(\rho^t)$. Thus, we have

$$E\Big(\sup_{\theta\in\Theta}\epsilon_t(\theta)\Big)=O(\rho^t).$$

and

$$E\Big(\sup_{\theta\in\Theta}|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)|\Big)\le\sum_{j=0}^{t-1}\rho^jO(\rho^{t-j})+\rho^tE|\lambda_0(\theta)-\widetilde{\lambda}_0(\theta)|=O(t\rho^t).$$

Choose an appropriate constant $\rho^*\in(\rho,1)$ such that $t\rho^t=O(\rho^{*t})$. Thus, (ii) holds.

By the Minkowski inequality, (iii) and (iv) can be proved similarly. □

Lemma 6. *If Assumptions 1-3 hold, then,*

$$\sup_{\theta\in\Theta}n^{-1}|L_n(\theta)-\widetilde{L}_n(\theta)|\to 0\quad a.s.\quad as\ n\to\infty.$$

PROOF. By Lemma 5, it follows that $\sup_{\theta\in\Theta}|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)|\to 0$ and $\sup_{\theta\in\Theta}|y_t\{\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\}|\to 0$ a.s. Then we have

$$\begin{aligned}
&\sup_{\theta\in\Theta}n^{-1}|L_n(\theta)-\widetilde{L}_n(\theta)|\\
&=\sup_{\theta\in\Theta}n^{-1}\Big|\sum_{t=1}^{n}y_t\{\log\lambda_t(\theta)-\log\widetilde{\lambda}_t(\theta)\}-\{\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\}\Big|\\
&\le n^{-1}\Big\{\sup_{\theta\in\Theta}\sum_{t=1}^{n}\Big(y_t\Big|\frac{\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)}{\inf_{\theta\in\Theta}\min\{\omega_1,\omega_2\}}\Big|+|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)|\Big)\Big\}\\
&\le n^{-1}\sum_{t=1}^{n}\Big\{\sup_{\theta\in\Theta}\Big|\frac{y_t\{\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)\}}{\inf_{\theta\in\Theta}\min\{\omega_1,\omega_2\}}\Big|\Big\}+n^{-1}\sum_{t=1}^{n}\Big(\sup_{\theta\in\Theta}|\lambda_t(\theta)-\widetilde{\lambda}_t(\theta)|\Big)\\
&\to 0\quad a.s.\quad \text{as } n\to\infty.
\end{aligned}$$

The proof is complete. □

Lemma 7. *For any* $\theta \in \Theta$*, define* $V_\eta(\theta) = V_\eta((\vartheta^T, \tau^T)^T) = \{\widetilde{\theta} \mid \widetilde{\vartheta} \in B(\vartheta, \eta)\}$*. If Assumptions 1-3 hold, then*

$$E\Big\{\sup_{\widetilde{\theta}\in V_\eta(\theta)} |\ell_t(\widetilde{\theta}) - \ell_t(\theta)|\Big\} \to 0 \quad as\ \eta \to 0.$$

Proof. Denote $a_{t-1} = (\omega_1 + \alpha_1 y_{t-1})R_t(\theta) + (\omega_2 + \alpha_2 y_{t-1})\{1 - R_t(\theta)\}$, $b_{t-1} = \beta_1 R_t(\theta) + \beta_2\{1 - R_t(\theta)\}$, respectively. Then $|a_t| \le C(1 + y_t)$, $|b_t| \le \rho < 1$, and

$$\begin{aligned}
\lambda_t(\theta) &= (\omega_1 + \alpha_1 y_{t-1})R_t(\theta) + (\omega_2 + \alpha_2 y_{t-1})(1 - R_t(\theta)) \\
&\quad + \{\beta_1 R_t(\theta) + \beta_2(1 - R_t(\theta))\}\lambda_{t-1}(\theta) \\
&= a_{t-1} + b_{t-1}\lambda_{t-1}(\theta) \\
&= \sum_{k=1}^{\infty}\prod_{j=1}^{k-1} b_{t-j}a_{t-k},
\end{aligned}$$

here the convention $\prod_{j=1}^{0}\cdot = 1$ is adopted. Thus, by an inequality $|\prod_{i=1}^n z_{1i} - \prod_{i=1}^n z_{2i}| \le \eta^{n-1}\sum_{i=1}^n |z_{1i} - z_{2i}|$ for $|z_{1i}|, |z_{2i}| \le \eta$, for $\widetilde{\theta} \in V_\eta(\theta)$, it follows that

$$\begin{aligned}
|\lambda_t(\theta) - \lambda_t(\widetilde{\theta})| &= \Big|\sum_{k=1}^{\infty}\prod_{j=1}^{k-1} b_{t-j}a_{t-k} - \sum_{k=1}^{\infty}\prod_{j=1}^{k-1} \widetilde{b}_{t-j}\widetilde{a}_{t-k}\Big| \\
&= \Big|\sum_{k=1}^{\infty}\Big\{\Big(\prod_{j=1}^{k-1} b_{t-j} - \prod_{j=1}^{k-1}\widetilde{b}_{t-j}\Big)a_{t-k} + \prod_{j=1}^{k-1}\widetilde{b}_{t-j}(a_{t-k} - \widetilde{a}_{t-k})\Big\}\Big| \\
&\le C\eta\sum_{k=1}^{\infty} k\rho^k(1 + y_{t-k}).
\end{aligned}$$

Thus,

$$
\begin{aligned}
E\Big\{\sup_{\widetilde{\theta}\in V_\eta(\theta)} |\ell_t(\widetilde{\theta}) - \ell_t(\theta)|\Big\} &\le \Big\|\frac{y_t}{\inf_{\theta\in\Theta}\min\{\omega_1,\omega_2\}} + 1\Big\|_2 \Big\|\sup_{\widetilde{\theta}\in V_\eta(\theta)} |\lambda_t(\theta) - \lambda_t(\widetilde{\theta})|\Big\|_2 \\
&\le C\eta \Big\|\frac{y_t}{\inf_{\theta\in\Theta}\min\{\omega_1,\omega_2\}} + 1\Big\|_2 \sum_{k=1}^{\infty} k\rho^k (1 + \|y_{t-k}\|_2) \\
&\to 0 \quad \text{as } \eta \to 0,
\end{aligned}
$$

where $\|X\|_2 = \{E(X^2)\}^{1/2}$. The proof is complete. □

Lemma 8. *If Assumptions 1-3 hold, then* $E\{\ell_t(\theta)\} \le E\{\ell_t(\theta_0)\}$ *for any* $\theta \in \Theta$ *and the equality holds if and only if* $\theta = \theta_0$.

Proof. By the Jensen inequality, a conditional argument gives that

$$
\begin{aligned}
E\{\ell_t(\theta) - \ell_t(\theta_0)\} &= E\Big\{E\Big(\log \frac{\phi(y_t \mid \lambda_t(\theta))}{\phi(y_t \mid \lambda_t)} \;\Big|\; \mathcal{F}_{t-1}\Big)\Big\} \\
&\le E\Big\{\log E\Big(\frac{\phi(y_t \mid \lambda_t(\theta))}{\phi(y_t \mid \lambda_t)} \;\Big|\; \mathcal{F}_{t-1}\Big)\Big\} \\
&= E(\log 1) = 0,
\end{aligned}
$$

where $\phi(\cdot \mid x)$ denotes the Poisson density function with mean x and the equality holds if and only if $\lambda_t(\theta) = \lambda_t$. Suppose $\theta^* \in \Theta$ satisfies $\lambda_t(\theta^*) = \lambda_t$.

Then

$$
\begin{aligned}
0 &= \lambda_t(\theta^*) - \lambda_t \\
&= (\omega_1^* + \alpha_1^* y_{t-1} + \beta_1^* \lambda_{t-1}(\theta^*)) R_t(\theta^*) + (\omega_2^* + \alpha_2^* y_{t-1} + \beta_2^* \lambda_{t-1}(\theta^*))(1 - R_t(\theta^*)) \\
&\quad - \{(\omega_{10} + \alpha_{10} y_{t-1} + \beta_{10} \lambda_{t-1}) R_t + (\omega_{20} + \alpha_{20} y_{t-1} + \beta_{20} \lambda_{t-1})(1 - R_t)\} \\
&= (\omega_1^* + \alpha_1^* y_{t-1} + \beta_1^* \lambda_{t-1}) R_t(\theta^*) \{R_t + (1 - R_t)\} \\
&\quad + (\omega_2^* + \alpha_2^* y_{t-1} + \beta_2^* \lambda_{t-1})(1 - R_t(\theta^*)) \{R_t + (1 - R_t)\} \\
&\quad - (\omega_{10} + \alpha_{10} y_{t-1} + \beta_{10} \lambda_{t-1}) R_t \{R_t(\theta^*) + (1 - R_t(\theta^*))\} \\
&\quad - (\omega_{20} + \alpha_{20} y_{t-1} + \beta_{20} \lambda_{t-1})(1 - R_t) \{R_t(\theta^*) + (1 - R_t(\theta^*))\} \\
&= \{(\omega_1^* + \alpha_1^* y_{t-1} + \beta_1^* \lambda_{t-1}) - (\omega_{10} + \alpha_{10} y_{t-1} + \beta_{10} \lambda_{t-1})\} R_t(\theta^*) R_t \\
&\quad + \{(\omega_1^* + \alpha_1^* y_{t-1} + \beta_1^* \lambda_{t-1}) - (\omega_{20} + \alpha_{20} y_{t-1} + \beta_{20} \lambda_{t-1})\} R_t(\theta^*)(1 - R_t) \\
&\quad + \{(\omega_2^* + \alpha_2^* y_{t-1} + \beta_2^* \lambda_{t-1}) - (\omega_{10} + \alpha_{10} y_{t-1} + \beta_{10} \lambda_{t-1})\}(1 - R_t(\theta^*)) R_t \\
&\quad + \{(\omega_2^* + \alpha_2^* y_{t-1} + \beta_2^* \lambda_{t-1}) - (\omega_{20} + \alpha_{20} y_{t-1} + \beta_{20} \lambda_{t-1})\}(1 - R_t(\theta^*))(1 - R_t).
\end{aligned}
$$

Thus, we have

$$
(\omega_1^* + \alpha_1^* y_{t-1} + \beta_1^* \lambda_{t-1}) - (\omega_{10} + \alpha_{10} y_{t-1} + \beta_{10} \lambda_{t-1}) = 0 \quad a.s.
$$

$$
(\omega_2^* + \alpha_2^* y_{t-1} + \beta_2^* \lambda_{t-1}) - (\omega_{20} + \alpha_{20} y_{t-1} + \beta_{20} \lambda_{t-1}) = 0 \quad a.s.
$$

Clearly, $\omega_1^* = \omega_{10}$, $\omega_2^* = \omega_{20}$. Then $(\alpha_1^* - \alpha_{10}) + (\beta_1^* - \beta_{10}) = 0$ and $(\alpha_2^* - \alpha_{20}) + (\beta_2^* - \beta_{20}) = 0$ by $\{(\alpha_1^* - \alpha_{10}) + (\beta_1^* - \beta_{10})\} E(\lambda_{t-1}) = 0$ and $\{(\alpha_2^* - \alpha_{20}) + (\beta_2^* - \beta_{20})\} E(\lambda_{t-1}) = 0$. Since $\mathcal{L}(y_{t-1} \mid \mathcal{F}_{t-2}) = \text{Poisson}(\lambda_{t-1})$, we have $\alpha_1^* = \alpha_{10}$, $\alpha_2^* = \alpha_{20}$, $\beta_1^* = \beta_{10}$ and $\beta_2^* = \beta_{20}$.

Further, by Assumption 1, we have $R_t(\theta^*) - R_t = 0$. It follows that

$$0 = R_t(\theta^*) - R_t = \{I(y_{t-1} \le s^*) - I(y_{t-1} \le s_0)\}R_{t-1} + \{I(y_{t-1} \le r^*) - I(y_{t-1} \le r_0)\}(1 - R_{t-1}).$$

By the orthogonality between R_{t-1} and $1 - R_{t-1}$, we can get

$$I(y_{t-1} \le s^*) - I(y_{t-1} \le s_0) = 0, \quad I(y_{t-1} \le r^*) - I(y_{t-1} \le r_0) = 0, \quad \text{a.s.}$$

Thus $r^* = r_0$ and $s^* = s_0$. Therefore, $\theta^* = \theta_0$. The proof is complete. □

Lemma 9. *If Assumptions 1-4 hold, then*

(i). $E\Big\{\sup_{\theta\in\Theta}\Big|\dfrac{\partial L(\delta,\theta)}{\partial\delta} - \dfrac{\partial \widetilde{L}(\delta,\theta)}{\partial\delta}\Big|\Big\} = O(1)$;

(ii). $E\Big\{\sup_{\theta\in\Theta}\Big|\dfrac{\partial^2 L(\delta,\theta)}{\partial^2\delta} - \dfrac{\partial^2 \widetilde{L}(\delta,\theta)}{\partial^2\delta}\Big|\Big\} = O(1), \quad \delta = 0, 1.$

Proof. It is sufficient to prove the case of $\delta = 0$. From the above proof, there exists a constant $\rho \in (0,1)$ such that for $k \in \mathbb{N}$, as $t \to \infty$,

$$E\big[\sup_{\theta\in\Theta}|\lambda_t^{\mathrm{b}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)|^k\big] = O(\rho^t), \quad E\big[\sup_{\theta\in\Theta}|\lambda_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{h}}(\theta)|^k\big] = O(\rho^t).$$

Applying the Cauchy inequality, we have

$$\begin{aligned}
&E\Big\{\sup_{\theta\in\Theta}\Big|\frac{\partial L(0,\theta)}{\partial\delta} - \frac{\partial \widetilde{L}(0,\theta)}{\partial\delta}\Big|\Big\} \\
&\le \sum_{t=1}^{n} E\Big[\sup_{\theta\in\Theta}\Big|\frac{\partial \lambda_t(0,\theta)}{\partial\delta} - \frac{\partial \widetilde{\lambda}_t(0,\theta)}{\partial\delta}\Big|\Big| -1 + \frac{y_t}{\lambda_t^{\mathrm{b}}(\theta)}\Big| + \Big|\frac{\partial \widetilde{\lambda}_t(0,\theta)}{\partial\delta}\Big|\Big|\frac{y_t}{\lambda_t^{\mathrm{b}}(\theta)} - \frac{y_t}{\widetilde{\lambda}_t^{\mathrm{b}}(\theta)}\Big|\Big] \\
&\le \sum_{t=1}^{n} E\Big[\sup_{\theta\in\Theta}\Big|\lambda_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \lambda_t^{\mathrm{b}}(\theta) + \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\Big|\Big| -1 + \frac{y_t}{\lambda_t^{\mathrm{b}}(\theta)}\Big| + \Big|\widetilde{\lambda}_t^{\mathrm{h}}(\theta) - \widetilde{\lambda}_t^{\mathrm{b}}(\theta)\Big|\Big|\frac{y_t\{\widetilde{\lambda}_t^{\mathrm{b}}(\theta) - \lambda_t^{\mathrm{b}}(\theta)\}}{\lambda_t^{\mathrm{b}}(\theta)\widetilde{\lambda}_t^{\mathrm{b}}(\theta)}\Big|\Big] \\
&\le \sum_{t=1}^{n} C\rho^t = O(1)
\end{aligned}$$

and

$$\begin{aligned}
&E\Big\{\sup_{\theta\in\Theta}\Big|\frac{\partial^2 L\,(0,\theta)}{\partial^2\delta}-\frac{\partial^2 \widetilde{L}\,(0,\theta)}{\partial^2\delta}\Big|\Big\}\\
&\le\sum_{t=1}^{n}E\Big[\sup_{\theta\in\Theta}\Big|\frac{y_t}{\{\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2}\{\widetilde{\lambda}_t^{\mathrm{h}}(\theta)-\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2-\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\{\lambda_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)\}^2\Big|\Big]\\
&\le\sum_{t=1}^{n}E\Big[\sup_{\theta\in\Theta}\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\Big|\{\lambda_t^{\mathrm{h}}(\theta)-\widetilde{\lambda}_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)+\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}\{\lambda_t^{\mathrm{h}}(\theta)+\widetilde{\lambda}_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)-\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}\Big|\\
&\quad+\Big|\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}-\frac{y_t}{\{\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2}\Big|\{\widetilde{\lambda}_t^{\mathrm{h}}(\theta)-\widetilde{\lambda}_t^{\mathrm{b}}(\theta)\}^2\Big]\le\sum_{t=1}^{n}C\rho^t=O(1).
\end{aligned}$$

Lemma 10. *If Assumptions 1-4 hold, then for* $\delta=0,1$*,*

$$E\Big\{\sup_{\theta\in\Theta}\frac{1}{n}\Big\|\frac{\partial^3 L\,(\delta,\theta)}{\partial\delta^2\partial\vartheta}\Big\|\Big\}=O(1).$$

Proof. Take $\delta=0$ as an example. It is not hard to show that

$$E\Big\{\frac{\partial\lambda_t(\theta)}{\partial\vartheta^i}\Big\}^p=O(1),\quad E\Big\{\frac{\partial^2\lambda_t(\theta)}{\partial\vartheta^i\partial\vartheta^j}\Big\}^p=O(1),\quad i,j=1,...,6,\quad p\in\mathbb{N},$$

by the Minkowski inequality. Thus, for $i=1,...,6$,

$$\begin{aligned}
&E\Big\{\sup_{\theta\in\Theta}\frac{1}{n}\frac{\partial^3 L\,(0,\theta)}{\partial\delta^2\partial\vartheta^i}\Big\}\\
&\le\frac{2}{n}\sum_{t=1}^{n}E\Big[\sup_{\theta\in\Theta}\Big|\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^3}\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i}\{\lambda_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)\}^2\\
&\qquad-\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\{\lambda_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)\}\Big\{\frac{\partial\lambda_t^{\mathrm{h}}(\theta)}{\partial\vartheta^i}-\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i}\Big\}\Big|\Big]\\
&\le\frac{1}{n}\sum_{t=1}^{n}C=O(1).
\end{aligned}$$

Lemma 11. *If Assumptions 1-4 hold, then for* $\delta=0,1$*,*

$$E\Big\{\sup_{\theta\in\Theta}\frac{1}{n}\Big\|\frac{\partial^3 L\,(\delta,\theta)}{\partial\delta\partial\vartheta\partial\vartheta^T}\Big\|\Big\}=O(1).$$

Proof. Take $\delta = 0$ as an example. For $i, j = 1, ..., 6$,

$$
\begin{aligned}
&E\Big\{\sup_{\theta\in\Theta}\frac{1}{n}\frac{\partial^3 L(0,\theta)}{\partial\delta\partial\vartheta^i\partial\vartheta^j}\Big\} \\
&\leq \frac{1}{n}\sum_{t=1}^{n} E\Big[\sup_{\theta\in\Theta}\Big|2\frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^3}\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i}\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^j}\{\lambda_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)\} - \frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\frac{\partial^2\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i\partial\vartheta^j}\{\lambda_t^{\mathrm{h}}(\theta)-\lambda_t^{\mathrm{b}}(\theta)\} \\
&\quad - \frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i}\Big\{\frac{\partial\lambda_t^{\mathrm{h}}(\theta)}{\partial\vartheta^j}-\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^j}\Big\} - \frac{y_t}{\{\lambda_t^{\mathrm{b}}(\theta)\}^2}\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^j}\Big\{\frac{\partial\lambda_t^{\mathrm{h}}(\theta)}{\partial\vartheta^i}-\frac{\partial\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i}\Big\} \\
&\quad + \Big(\frac{y_t}{\lambda_t^{\mathrm{b}}(\theta)}-1\Big)\Big\{\frac{\partial^2\lambda_t^{\mathrm{h}}(\theta)}{\partial\vartheta^i\partial\vartheta^j}-\frac{\partial^2\lambda_t^{\mathrm{b}}(\theta)}{\partial\vartheta^i\partial\vartheta^j}\Big\}\Big|\Big] \leq \frac{1}{n}\sum_{t=1}^{n} C = O(1).
\end{aligned}
$$

Bibliography

Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *J. Amer. Statist. Assoc. 104* (488), 1430–1439.

Wang, C., H. Liu, J.-f. Yao, R. A. Davis, and W. K. Li (2014). Self-excited threshold Poisson autoregression. *J. Amer. Statist. Assoc. 109* (506), 777–787.