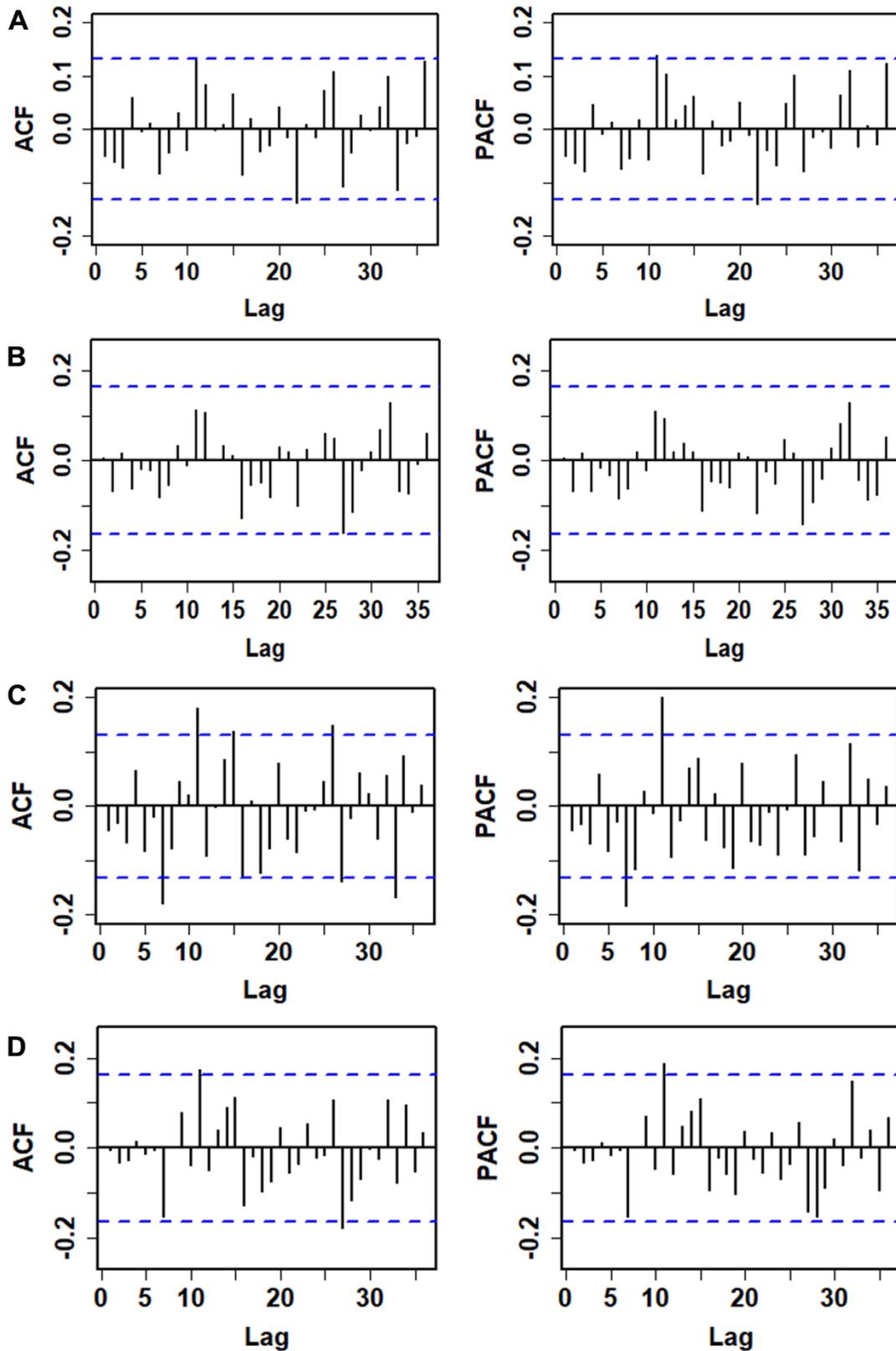
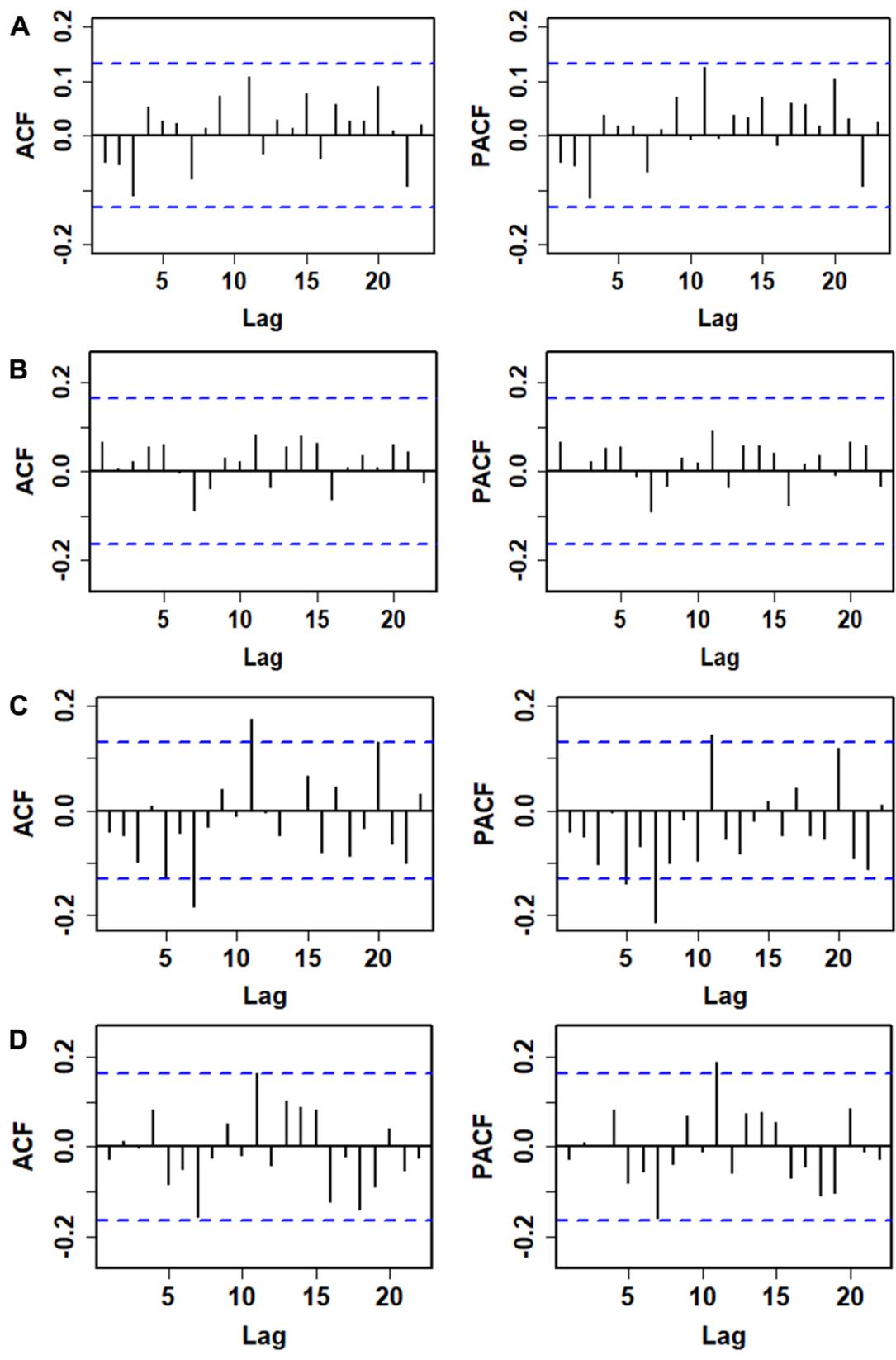


Supplementary Figure 1 Autocorrelation function and partial autocorrelation function plot for the seasonally differenced Hepatitis B incidence series. It is apparent that significant spikes at the lags of 11 and 12 in the autocorrelation function (ACF) analysis, and the lag of 12 in the partial autocorrelation function (PACF) analysis, suggesting that the maximum order of seasonal autoregressive was 1 and the maximum orders of seasonal moving average were 2. The significant first three spikes in the PACF analysis indicated the maximum orders of autoregressive being 3, and the trailing form in the ACF analysis showed the maximum orders of moving average being 1. ACF: Autocorrelation function; PACF: Partial autocorrelation function.



Supplementary Figure 2 Autocorrelation function and partial autocorrelation function for the resulting errors of hepatitis B and C from the optimal seasonal autoregressive integrated moving average. A: Autocorrelation

function (ACF) and partial autocorrelation function (PACF) plots of the 12-step ahead forecasts for hepatitis B; B: ACF and PACF plots of the 90-step ahead forecasts for hepatitis B; C: ACF and PACF plots of the 12-step ahead forecasts for hepatitis C; D: ACF and PACF plots of the 90-step ahead forecasts for hepatitis C. As depicted in these plots, most spikes fell within the significant limits except for few ones in Supplementary Figure 2A-D (This is also reasonable because some high-order lags easily touch the significance bounds by chance alone), being indicative of a white noise series of the resulting series and meaning that the selected model can adequately fit the target series. ACF: Autocorrelation function; PACF: Partial autocorrelation function.



Supplementary Figure 3 Autocorrelation function and partial autocorrelation function analysis for the resulting errors of hepatitis B and C from the

optimal seasonal autoregressive fractionally integrated moving average. A: Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the 12-step ahead forecasts for hepatitis B; B: ACF and PACF plots of the 90-step ahead forecasts for hepatitis B; C: ACF and PACF plots of the 12-step ahead forecasts for hepatitis C; D: ACF and PACF plots of the 90-step ahead forecasts for hepatitis C. As depicted in these plots, most spikes fell within the significant limits except for few ones in Supplementary Figure 3C and D (This is also reasonable because some high-order lags easily touch the significance bounds by chance alone), being indicative of a white noise series of the resulting series and meaning that the selected model can adequately fit the target series. ACF: Autocorrelation function; PACF: Partial autocorrelation function.

Supplementary Table 1 Selected seasonal autoregressive integrated moving average candidates with their Akaike's information criterion, corrected Akaike's information criterion, Bayesian information criterion, and log-likelihood

Models	AIC	CAIC	BIC	LL
SARIMA (3, 0, 0) (0, 1, 2) ₁₂	4363.97	4364.53	4387.4	-2174.99
SARIMA (1, 0, 2) (2, 0, 0) ₁₂	4643.83	4644.5	4671.05	-2319.91
SARIMA (1, 0, 2) (2, 1, 0) ₁₂	4367.62	4368.17	4391.05	-2173.81
SARIMA (3, 0, 0) (1, 1, 0) ₁₂	4399.21	4399.62	4419.29	-2193.6
SARIMA (3, 0, 0) (1, 1, 1) ₁₂	4365.38	4365.94	4388.81	-2175.69
SARIMA (3, 0, 0) (0, 1, 1) ₁₂	4367.68	4368.09	4387.76	-2177.84
SARIMA (2, 0, 1) (0, 1, 1) ₁₂	4365.34	4365.75	4385.42	-2176.67
SARIMA (1, 0, 1) (0, 1, 1) ₁₂	4366.41	4366.71	4383.15	-2178.21

SARIMA: Seasonal autoregressive integrated moving average; AIC: Akaike's information criterion; CAIC: Corrected Akaike's information criterion; BIC: Bayesian information criterion; LL: Log-likelihood.

Supplementary Table 2 The resulting eight modes for the 90 holdout

forecasts of hepatitis B under the best seasonal autoregressive fractionally integrated moving average

Modes	AIC	BIC	LL
Mode 1	2610.359	2637.087	-1296.18
Mode 2	2610.748	2637.477	-1296.37
Mode 3	2614.732	2641.46	-1298.37
Mode 4	2622.996	2649.724	-1302.5
Mode 5	2628.471	2655.199	-1305.24
Mode 6	2636.169	2662.898	-1309.08
Mode 7	2656.232	2682.96	-1319.12
Mode 8	2670.183	2696.911	-1326.09

AIC: Akaike's information criterion; BIC: Bayesian information criterion; LL: Log-likelihood.

Supplementary Table 3 The resulting 21 modes for the 12 holdout forecasts of hepatitis C under the best seasonal autoregressive fractionally integrated moving average

Modes	AIC	BIC	LL
Mode 1	3257.009	3287.633	-1619.5
Mode 2	3270.839	3301.463	-1626.42
Mode 3	3279.88	3310.504	-1630.94
Mode 4	3280.849	3311.473	-1631.42
Mode 5	3285.044	3315.668	-1633.52
Mode 6	3290.815	3321.44	-1636.41
Mode 7	3291.692	3322.317	-1636.85
Mode 8	3301.947	3332.571	-1641.97
Mode 9	3305.475	3336.099	-1643.74
Mode 10	3309.404	3340.028	-1645.7
Mode 11	3309.407	3340.031	-1645.7
Mode 12	3309.411	3340.035	-1645.71
Mode 13	3309.415	3340.04	-1645.71

Mode 14	3309.421	3340.045	-1645.71
Mode 15	3309.422	3340.046	-1645.71
Mode 16	3309.424	3340.048	-1645.71
Mode 17	3309.425	3340.049	-1645.71
Mode 18	3309.429	3340.053	-1645.71
Mode 19	3309.43	3340.055	-1645.72
Mode 20	3309.431	3340.055	-1645.72
Mode 21	3338.489	3369.113	-1660.24

AIC: Akaike's information criterion; BIC: Bayesian information criterion; LL: Log-likelihood.

Supplementary Table 4 The resulting 20 modes for the 90 holdout forecasts of hepatitis C under the best seasonal autoregressive fractionally integrated moving average

Modes	AIC	BIC	LL
Mode 1	2036.609	2063.338	-1009.3
Mode 2	2037.943	2064.671	-1009.97
Mode 3	2038.883	2065.612	-1010.44
Mode 4	2041.959	2068.687	-1011.98
Mode 5	2042.638	2069.367	-1012.32
Mode 6	2042.73	2069.458	-1012.36
Mode 7	2043.397	2070.125	-1012.7
Mode 8	2046.526	2073.255	-1014.26
Mode 9	2047.774	2074.502	-1014.89
Mode 10	2051.938	2078.666	-1016.97
Mode 11	2053.813	2080.542	-1017.91
Mode 12	2059.945	2086.673	-1020.97
Mode 13	2062.605	2089.333	-1022.3
Mode 14	2065.604	2092.332	-1023.8
Mode 15	2109.291	2136.019	-1045.65
Mode 16	2110.704	2137.432	-1046.35

Mode 17	2114.203	2140.931	-1048.1
Mode 18	2120.411	2147.139	-1051.21
Mode 19	2123.154	2149.882	-1052.58
Mode 20	2138.137	2164.865	-1060.07

AIC: Akaike's information criterion; BIC: Bayesian information criterion; LL: Log-likelihood.

Supplementary Table 5 Comparison of the forecasting ability under the seasonal autoregressive integrated moving average and the seasonal autoregressive fractionally integrated moving average model constructed using the data from different age groups between January 2004 and December 2019

Metrics	Hepatitis B		Hepatitis C	
	SARIMA	SARFIMA	SARIMA	SARFIMA
0-14 years old ¹				
MAD	125.336	65.350	35.972	35.940
MAPE	0.200	0.105	0.348	0.347
RMSE	143.401	84.209	43.361	43.284
MER	0.190	0.099	0.282	0.282
RMSPE	0.235	0.135	0.467	0.466
15-64 years old ¹				
MAD	5143.208	4383.085	866.012	571.380
MAPE	0.072	0.060	0.064	0.043
RMSE	5874.019	5450.755	1056.119	781.632
MER	7.802	6.649	6.789	4.479
RMSPE	0.082	0.072	0.079	0.060
65 and above years old ¹				
MAD	1798.493	1653.961	697.902	650.415
MAPE	0.157	0.142	0.162	0.151
RMSE	1897.415	1797.900	757.843	707.928
MER	2.728	2.509	5.471	5.099

RMSPE 0.164 0.150 0.172 0.160

¹In Hepatitis B forecasting, the best SARIMA models were SARIMA (3, 1, 0) (0, 1, 2)₁₂ for the 0-14 years old, SARIMA (0, 1, 1) (2, 1, 1)₁₂ for the 15-64 years old, and SARIMA (2, 1, 1) (2, 1, 0)₁₂ for the 65 and above years old; the best SARFIMA models were SARFIMA (3, -0.232, 0) (0, 0.048, 2)₁₂ for 0-14 years old, SARFIMA (0, 0.484, 1) (2, -0.185, 1)₁₂ for 15-64 years old, and SARFIMA (2, -0.093, 1) (2, -0.437, 0)₁₂ for the 65 and above years old. In Hepatitis C forecasting, the best SARIMA models were SARIMA (1, 1, 1) (2, 0, 0)₁₂ for the 0-14 years old, SARIMA (3, 1, 0) (2, 1, 0)₁₂ for the 15-64 years old, and SARIMA (1, 1, 0) (2, 0, 2)₁₂ for the 65 and above years old; the best SARFIMA models were SARFIMA (1, -0.33, 0) (2, 0, 0)₁₂ for 0-14 years old, SARFIMA (3, -0.142, 0) (2, -0.203, 0)₁₂ for 15-64 years old, and SARFIMA (1, 0, 1) (2, -0.072, 2)₁₂ for the 65 and above years old. SARIMA: Seasonal autoregressive integrated moving average; SARFIMA: Seasonal autoregressive fractionally integrated moving average; MAD: Mean absolute deviation; MAPE: Mean absolute percentage error; RMSE: Root mean square error; MER: Mean error rate; RMSPE: Root mean square percentage error.

Supplementary Table 6 Comparison of the forecasting ability under the seasonal autoregressive integrated moving average and the seasonal autoregressive fractionally integrated moving average

Metrics	Hepatitis B		Hepatitis C	
	SARIMA (0, 1, 1) (2, 0, 2) ₁₂	SARFIMA (0, 0.499, 1) (2, 0.392, 2) ₁₂	SARIMA (0, 1, 1) (2, 0, 0) ₁₂	SARFIMA (1, 0, 1) (2, 0.365, 0) ₁₂
MAD	14780.282	14578.232	3282.314	3223.797
MAPE	0.177	0.174	0.2035	0.2034
RMSE	20172.797	19011.714	4055.280	3994.377
MER	0.145	0.143	0.174	0.171
RMSPE	0.277	0.268	0.284	0.289

Here the SARIMA and SARFIMA were developed by considering the habits or cultural patterns (infection risk) in the period of spring celebrations (the spring celebrations fall within January and February in China per year, thus a new variable was created, which was coded as "1" in January and February per year, while which was coded as "0" in other months per year). SARIMA: Seasonal autoregressive integrated moving average; SARFIMA: Seasonal autoregressive fractionally integrated moving average; MAD: Mean absolute deviation; MAPE: Mean absolute percentage error; RMSE: Root mean square error; MER: Mean error rate; RMSPE: Root mean square percentage error.