Appendix to Nowcasting Reported COVID-19 Hospitalizations Using De-Identified, Aggregated Medical Insurance Claims Data

Xueda Shen[1] Aaron Rumack[2] Bryan Wilder[2] Ryan J. Tibshirani[1]

[1]University of California, Berkeley [2]Carnegie Mellon University

A More details on claims signals

For the outpatient signal, any claim that has a primary ICD-10 code of U07.1, B97.21, or B97.29 is counted as confirmed COVID. For the inpatient signal, any claim with a primary ICD-10 code of U07.1, U07.2, B97.29, J12.81, Z03.818, B34.2, or J12.89 is counted as COVID-associated.

These definitions are consistent with the definitions of analogous outpatient and inpatient signals in the Delphi Epidata API. However, the signals that we use in our analysis differ from those in the Delphi Epidata API, at the time that this paper was written, in a way that relates to smoothing: the signals in the API are smoothed with a more sophisticated smoothing approach (which also explicitly adjusts for weekday/weekend differences), whereas the signals in this paper are smoothed via 7-day pooling (which implicitly accounts for weekday/weekend differences).

B Scenario 1: further investigation for NY

We present further analysis as to why state-level model for NY produces negative nowcasts over a portion of the month of June 2021, in scenario 1 (recall Figure 5 in the main paper). Figure 1 displays two versions of the outpatient signal over the month of May 2021, corresponding to issue dates of June 7 and June 8. We can see that a systematic upward revision of all signal values occurs on June 8. The nowcasting model, fit by training on data through the end of May (which does not see the revision on June 8), places a sizeable negative weight on the largest lag of the outpatient feature; once the upward revision occurs on June 8, the nowcasts on that and subsequent days display a significant downward trend. This is confirmed by Figure 2, which decomposes the nowcasts made in June into the contributions from each feature (simply the coefficient times the feature value). After June 7, we can clearly see that `out_20`, the largest lag of the outpatient figure, is responsible for driving the downward trend.

C Scenario 1: full set of backcasts

Figures 3–19 display backcasts at lag 0 (i.e., nowcasts), 5, and 10, for all 50 US states and DC, made by the mixed model in scenario 1, the monthly-update period. The format follows Figure 4 in the main paper.

D Scenario 2: full set of backcasts

Figures 20–36 display backcasts at lag 0 (i.e., nowcasts), 5, and 10, for all 50 US states and DC, made by the mixed model in scenario 2, the no-update period. The format follows Figure 7 in the main paper.

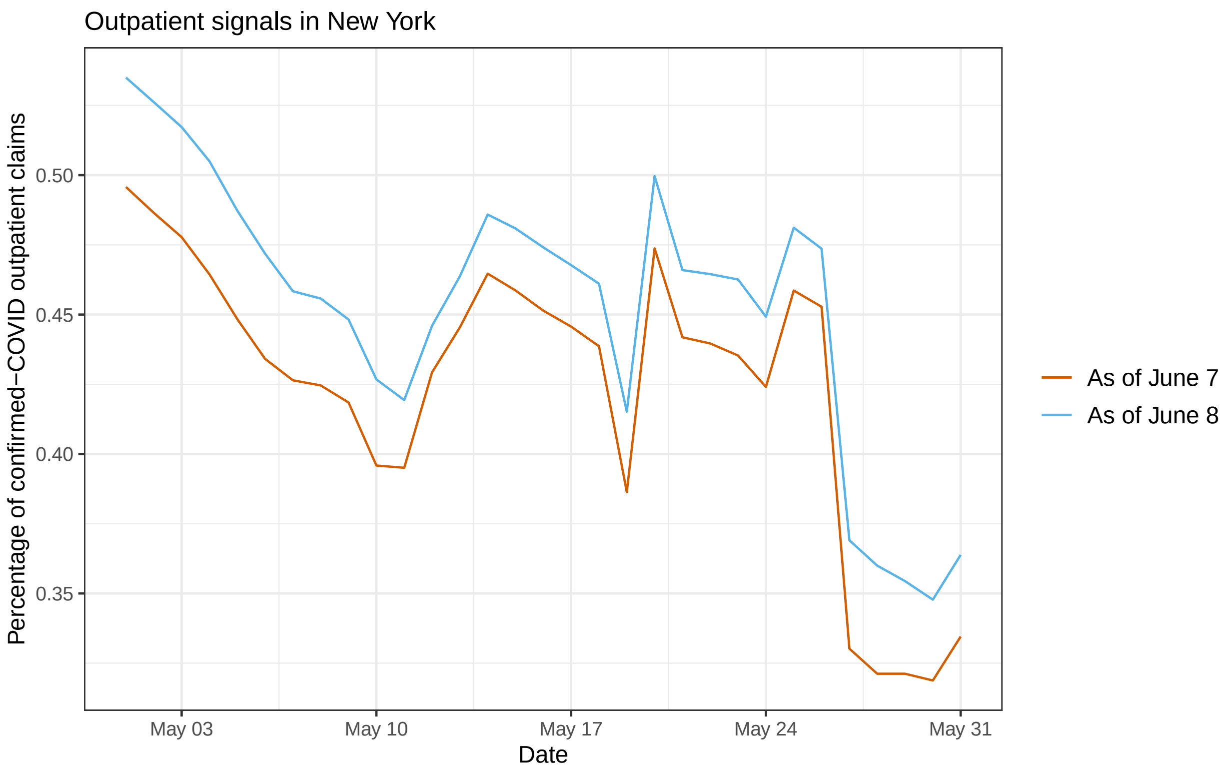

Figure 1: Two versions of the outpatient signal during May 2021, as of June 7 and June 8.

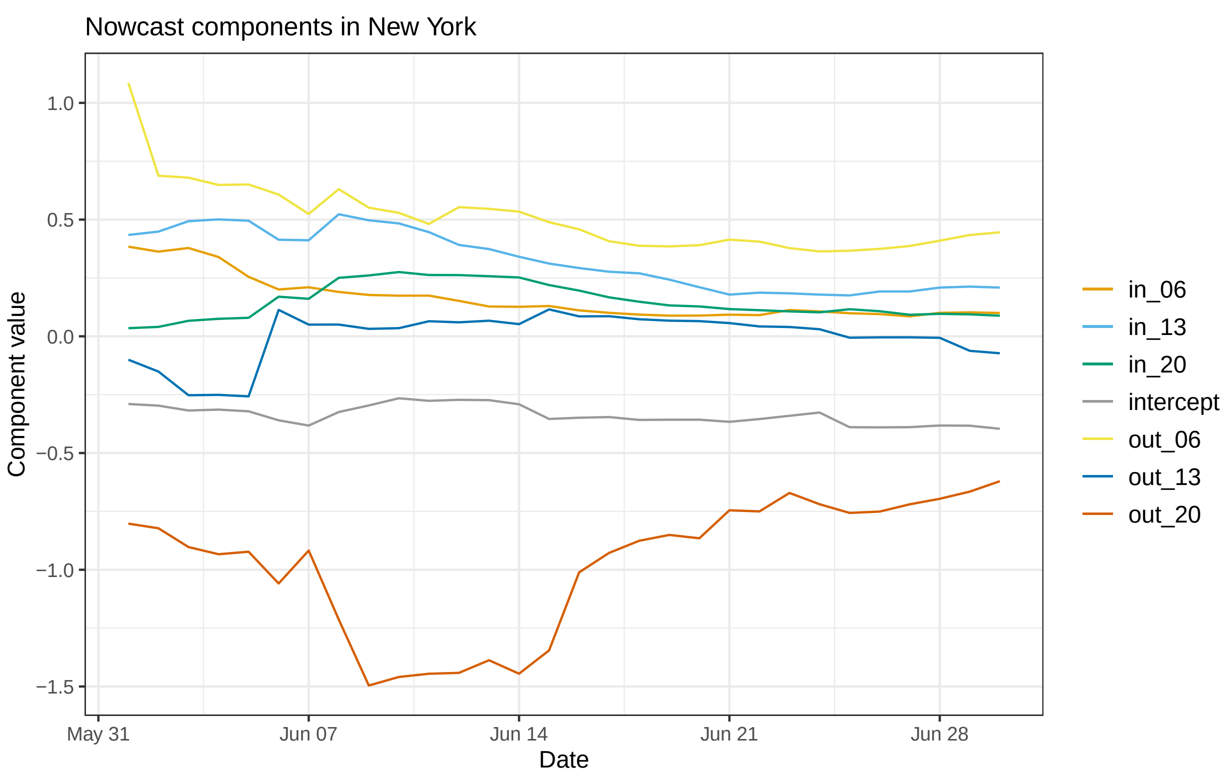

Figure 2: Contributions of each feature to the nowcasts during June 2021.

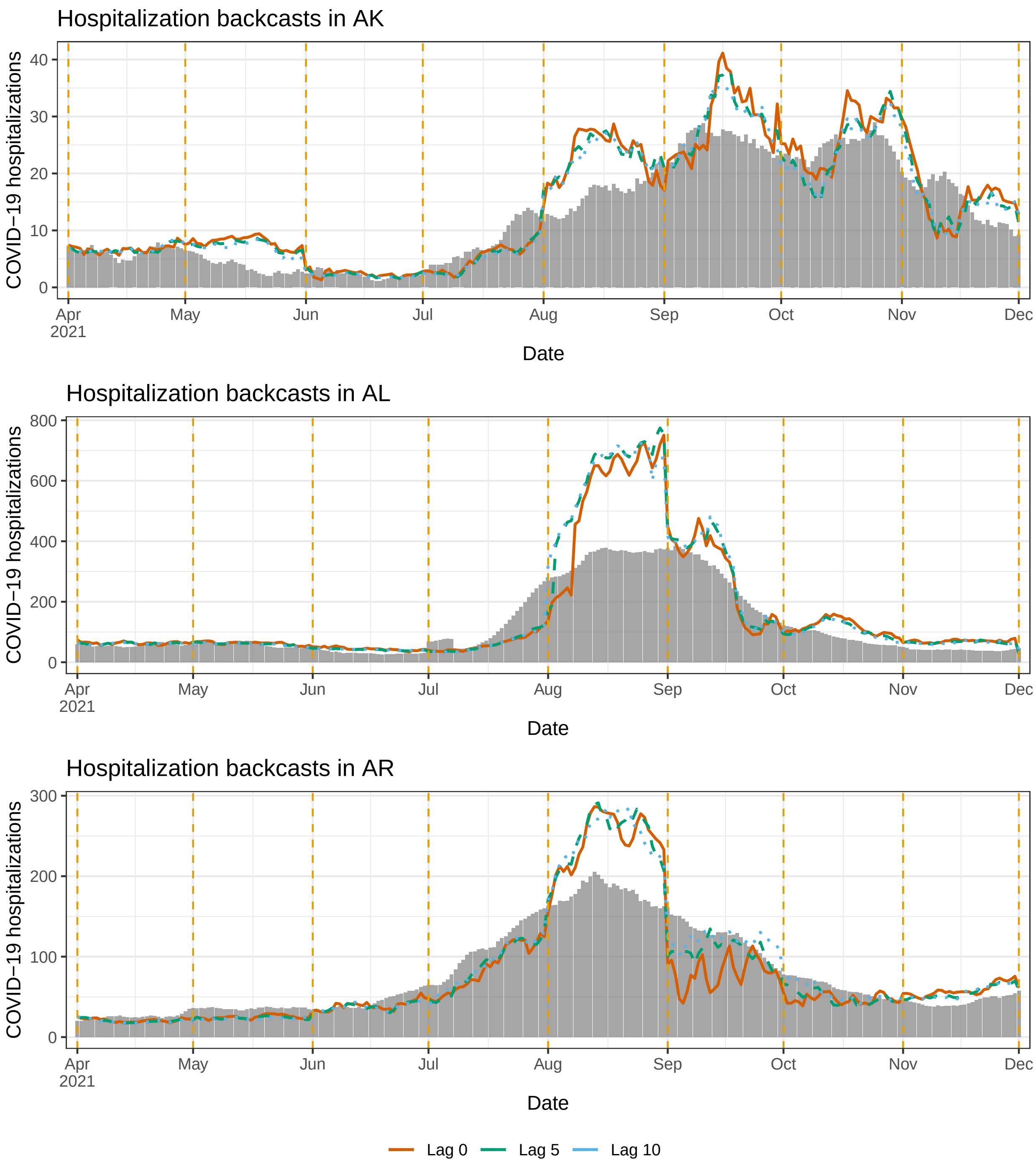

Figure 3: Backcasts from the mixed model in scenario 1, for AL, AK, AZ.

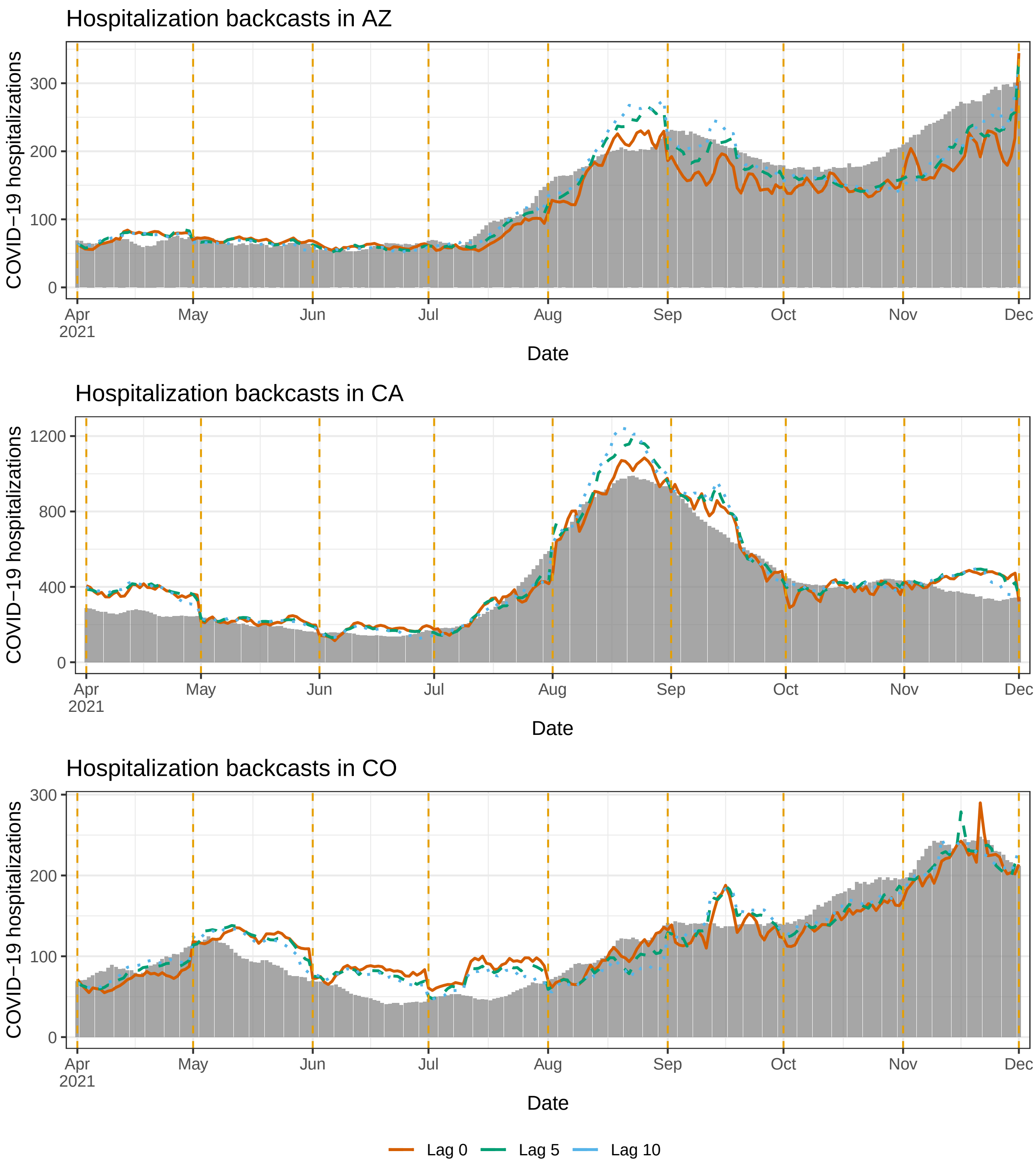

Figure 4: Backcasts from the mixed model in scenario 1, for AR, CA, CO.

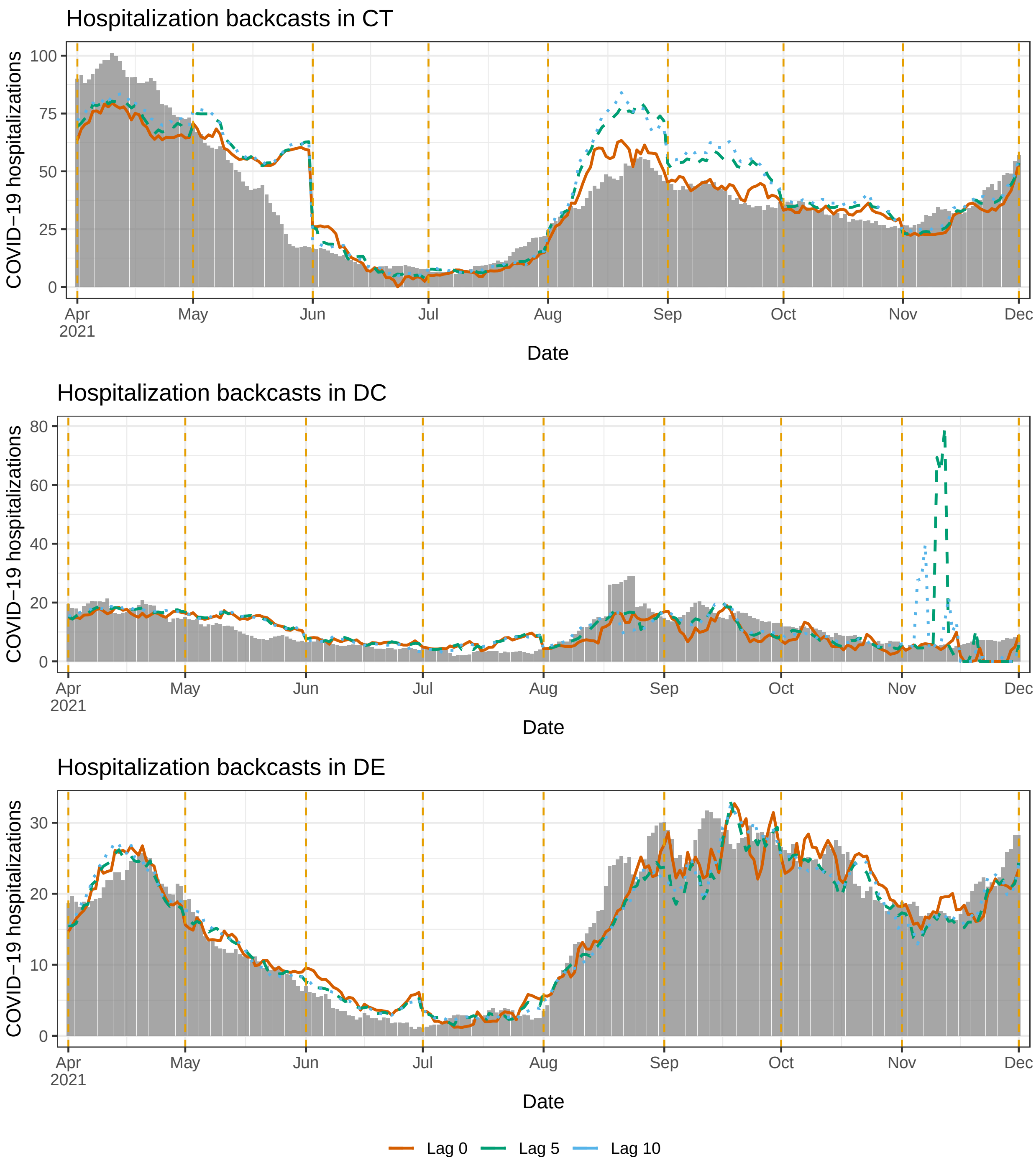

Figure 5: Backcasts from the mixed model in scenario 1, for CT, DC, DE.

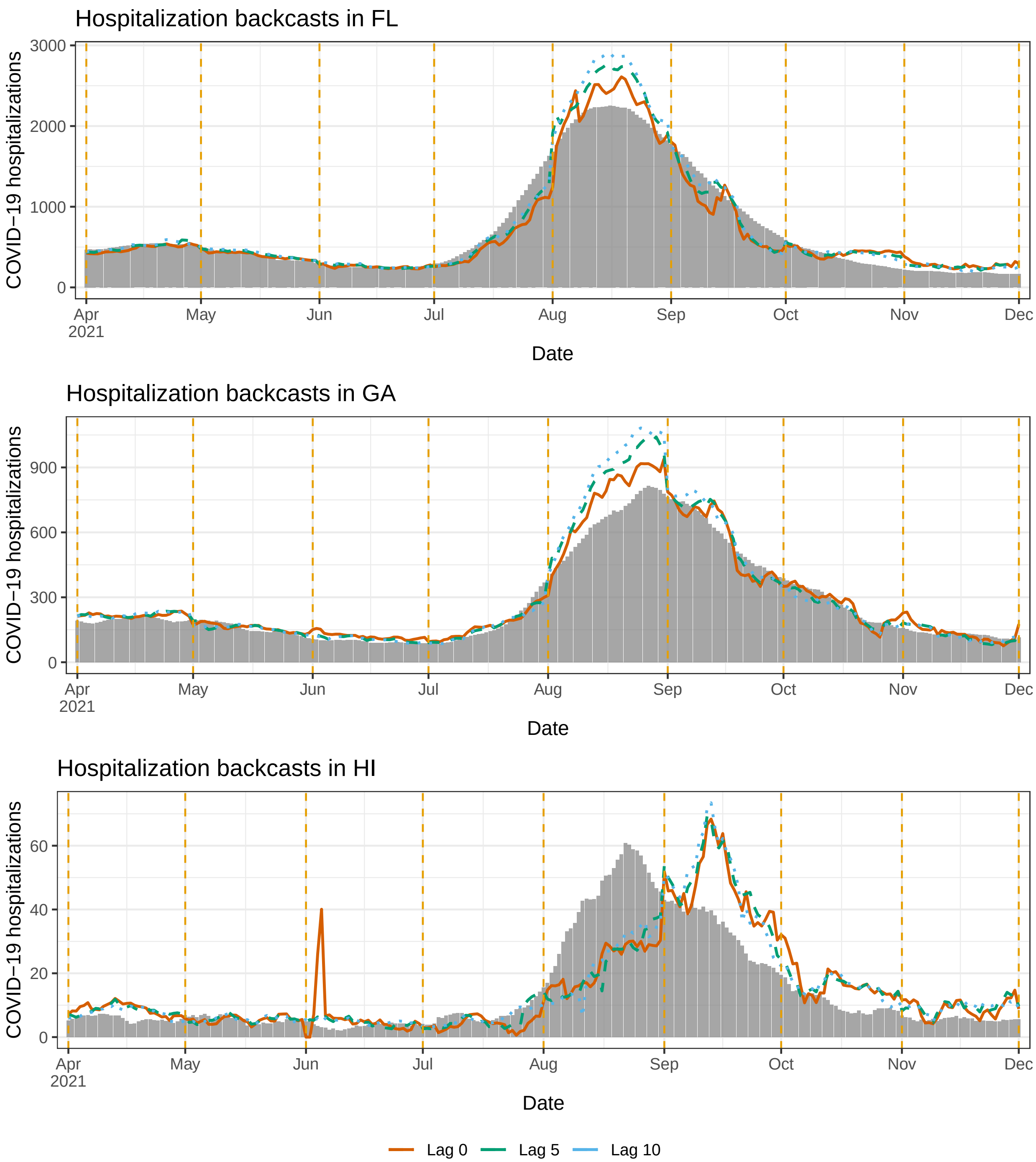

Figure 6: Backcasts from the mixed model in scenario 1, for FL, GA, HI.

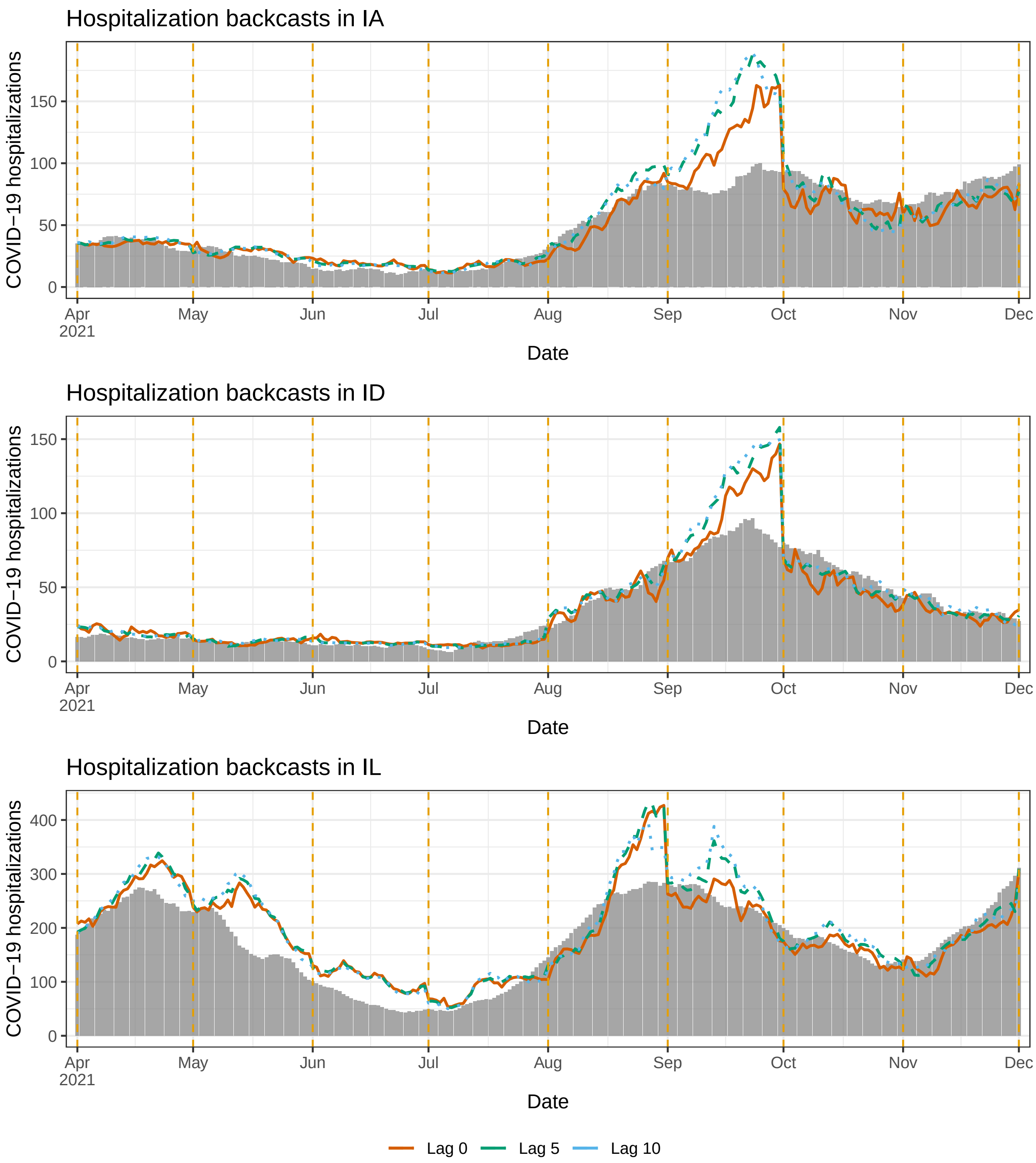

Figure 7: Backcasts from the mixed model in scenario 1, for IA, ID, IL.

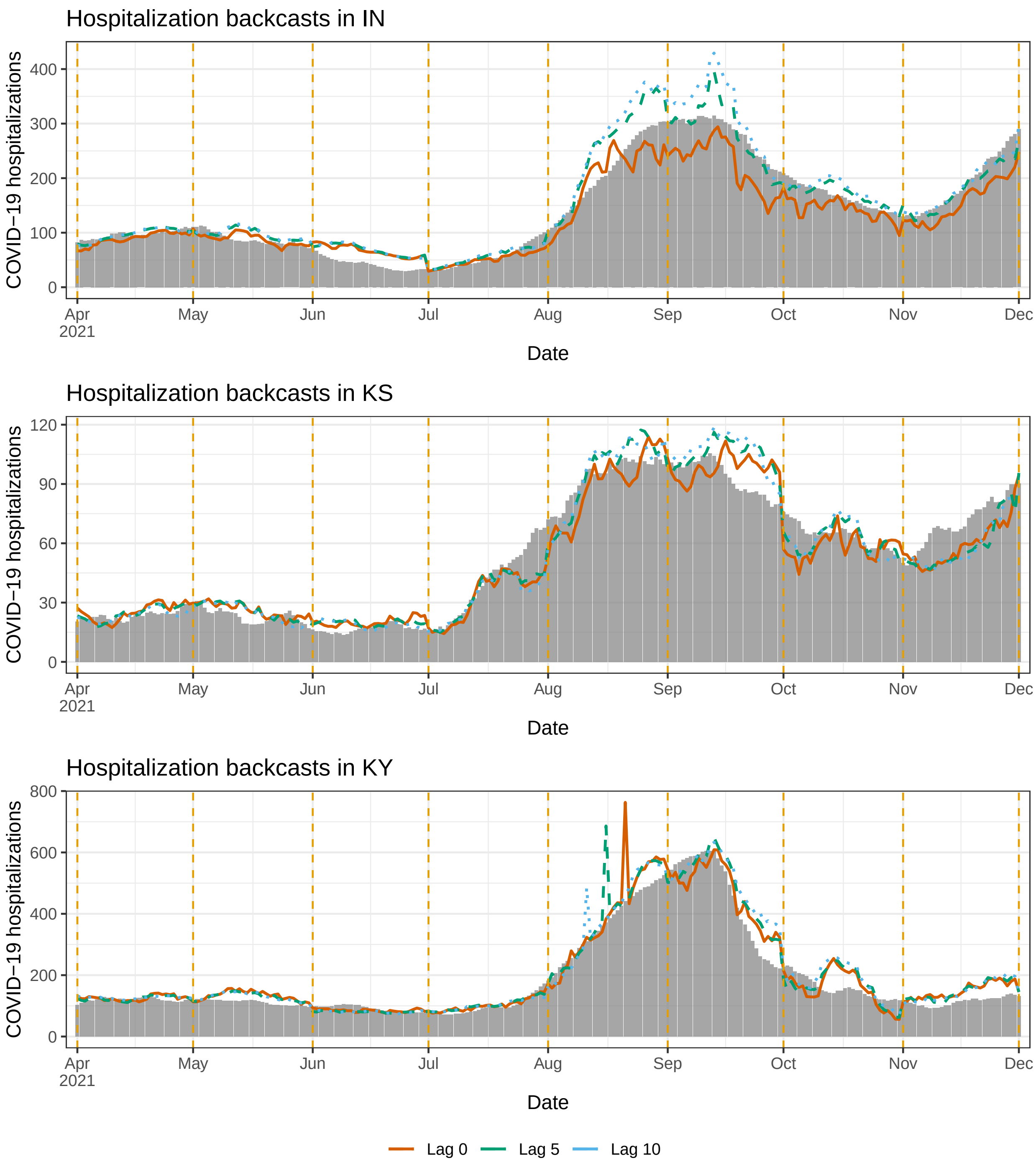

Figure 8: Backcasts from the mixed model in scenario 1, for IN, KS, KY.

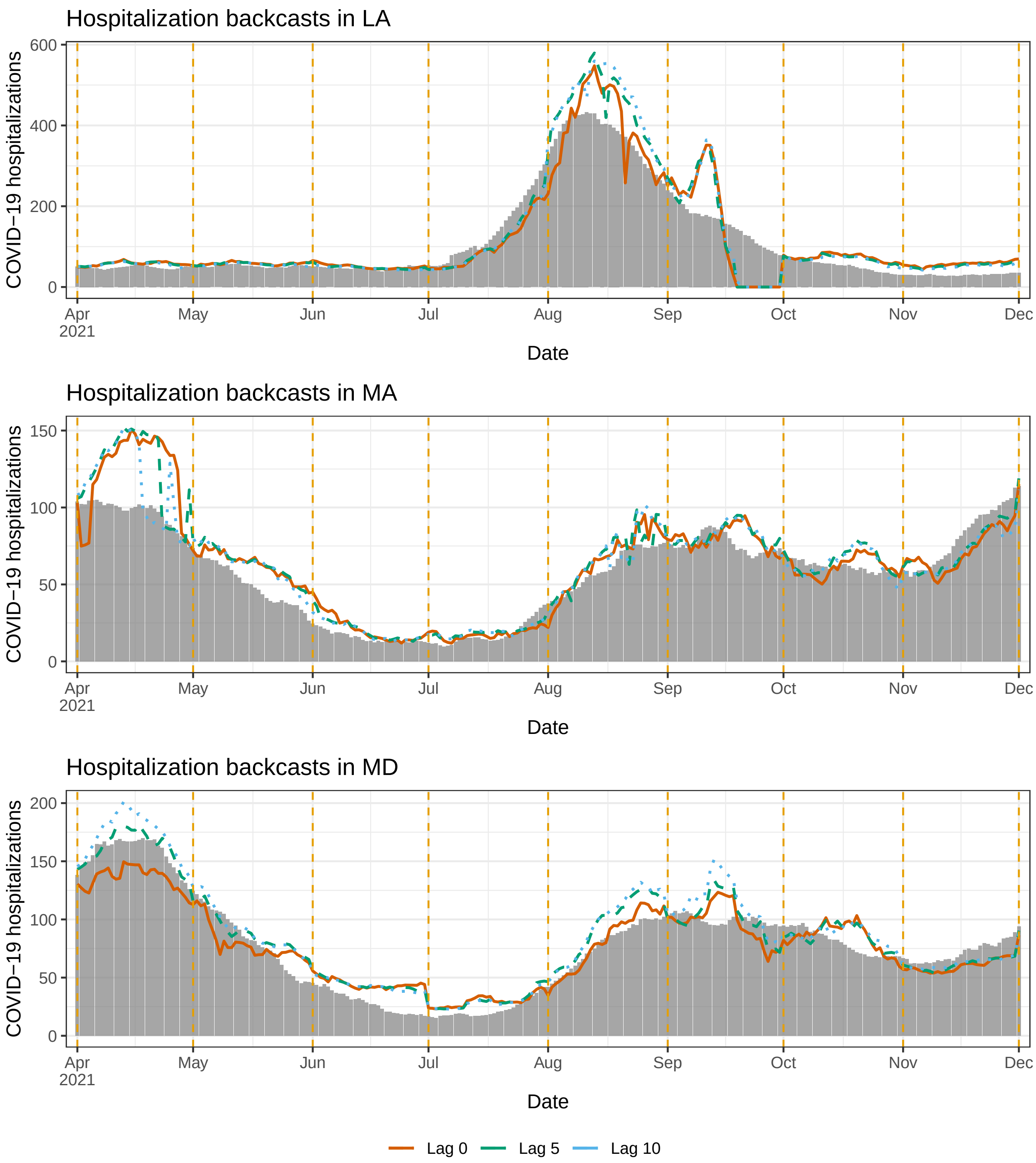

Figure 9: Backcasts from the mixed model in scenario 1, for LA, MA, MD.

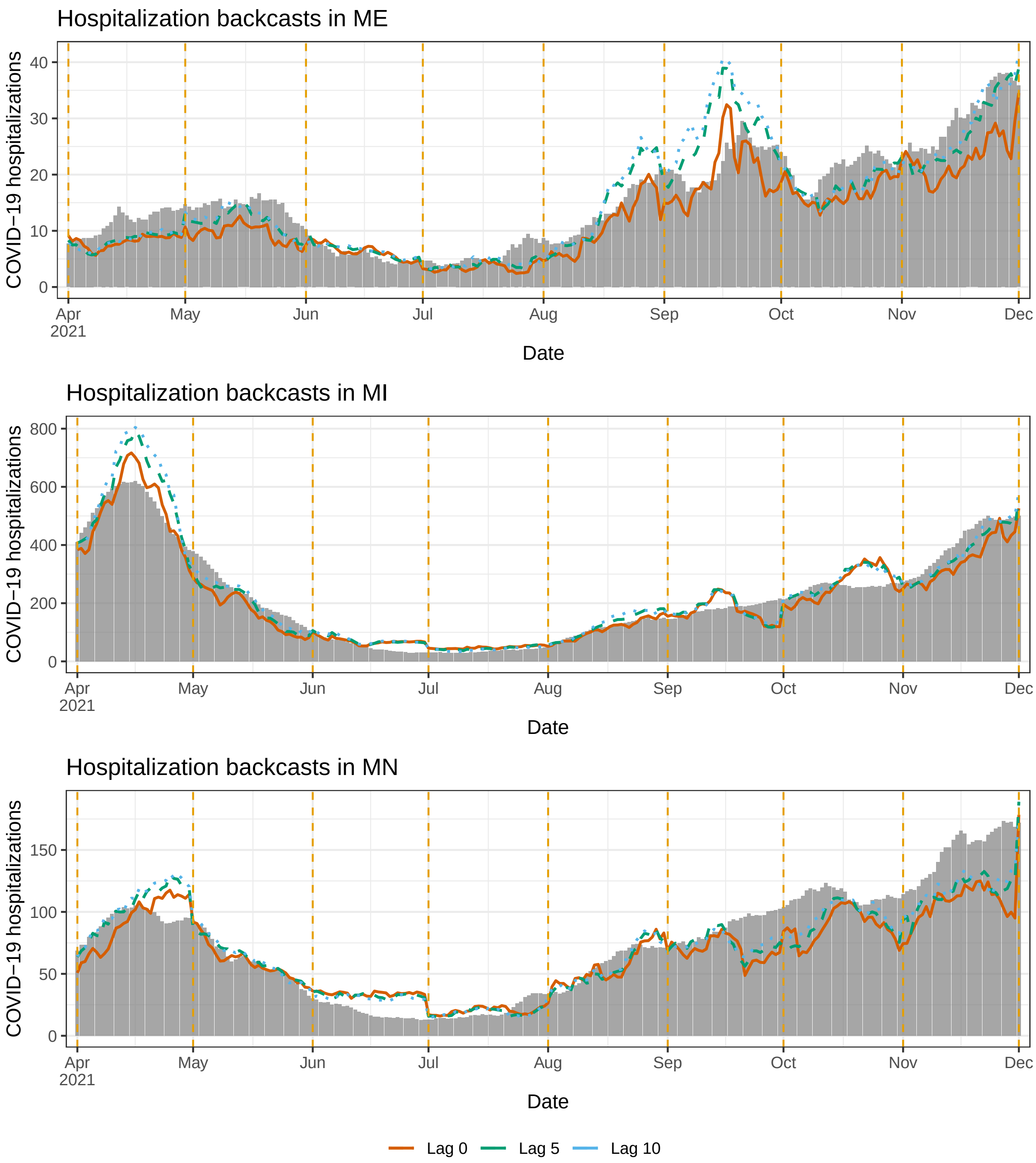

Figure 10: Backcasts from the mixed model in scenario 1, for ME, MI, MN.

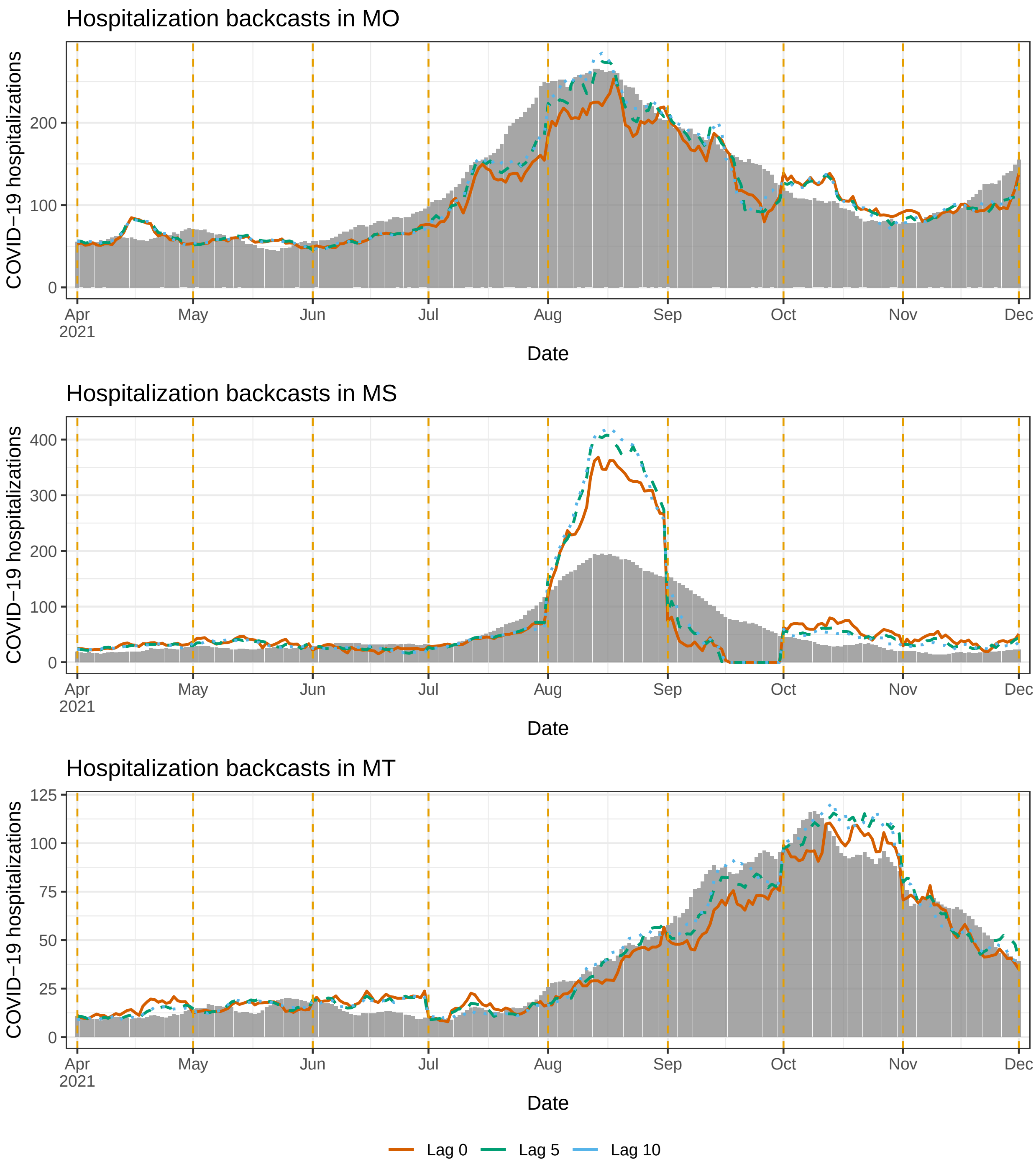

Figure 11: Backcasts from the mixed model in scenario 1, for MO, MS, MT.

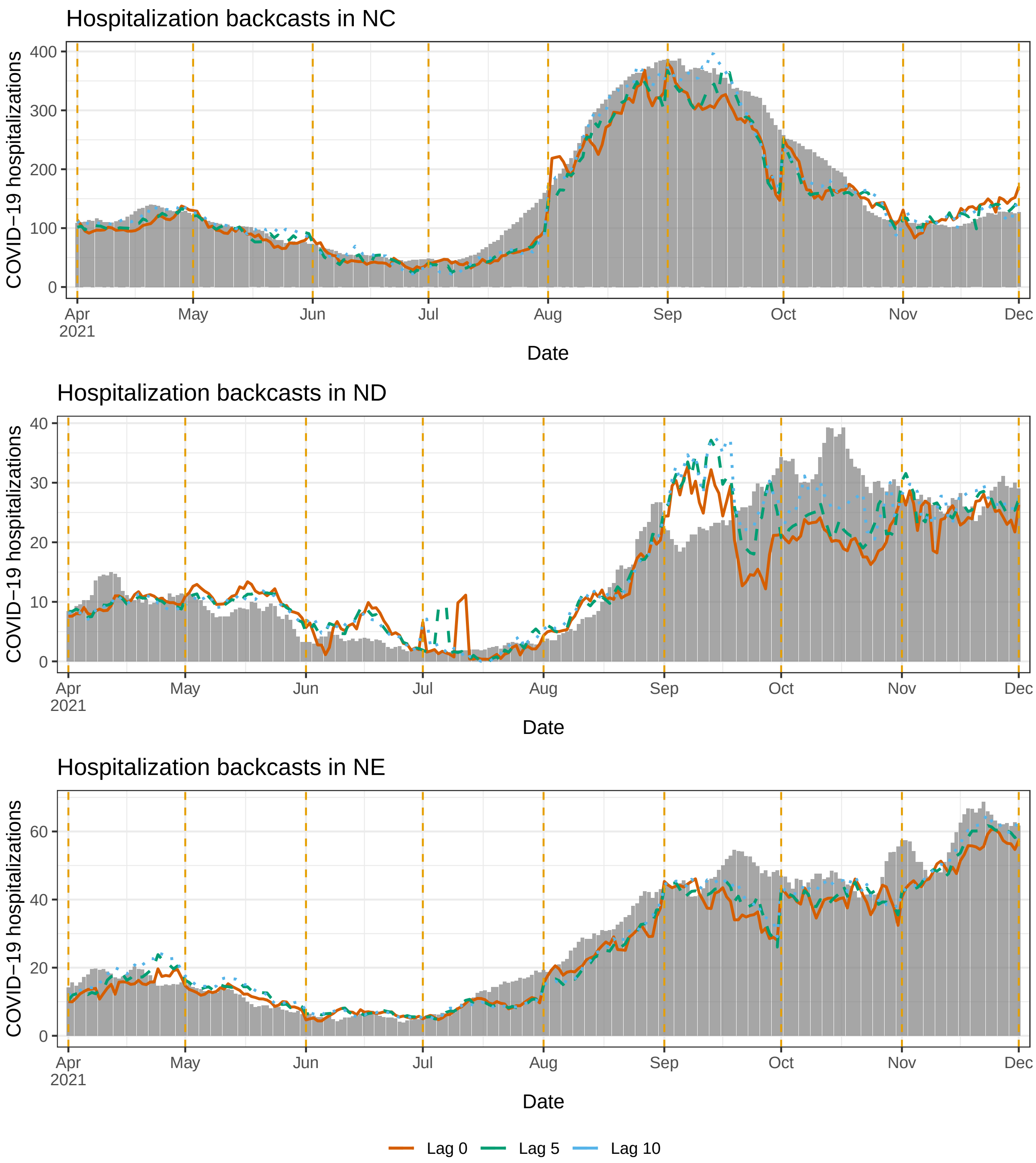

Figure 12: Backcasts from the mixed model in scenario 1, for NC, ND, NE.

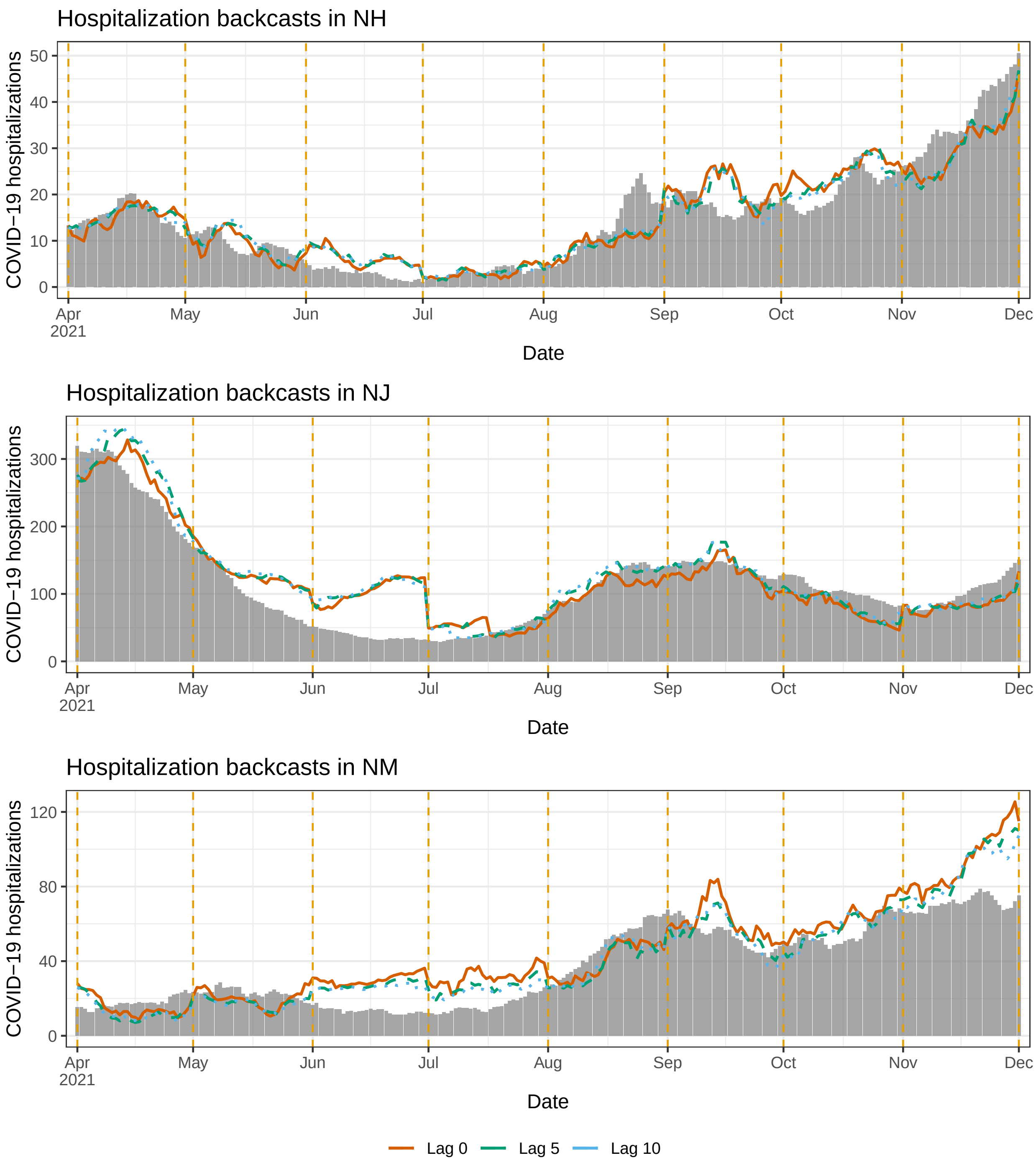

Figure 13: Backcasts from the mixed model in scenario 1, for NH, NJ, NM.

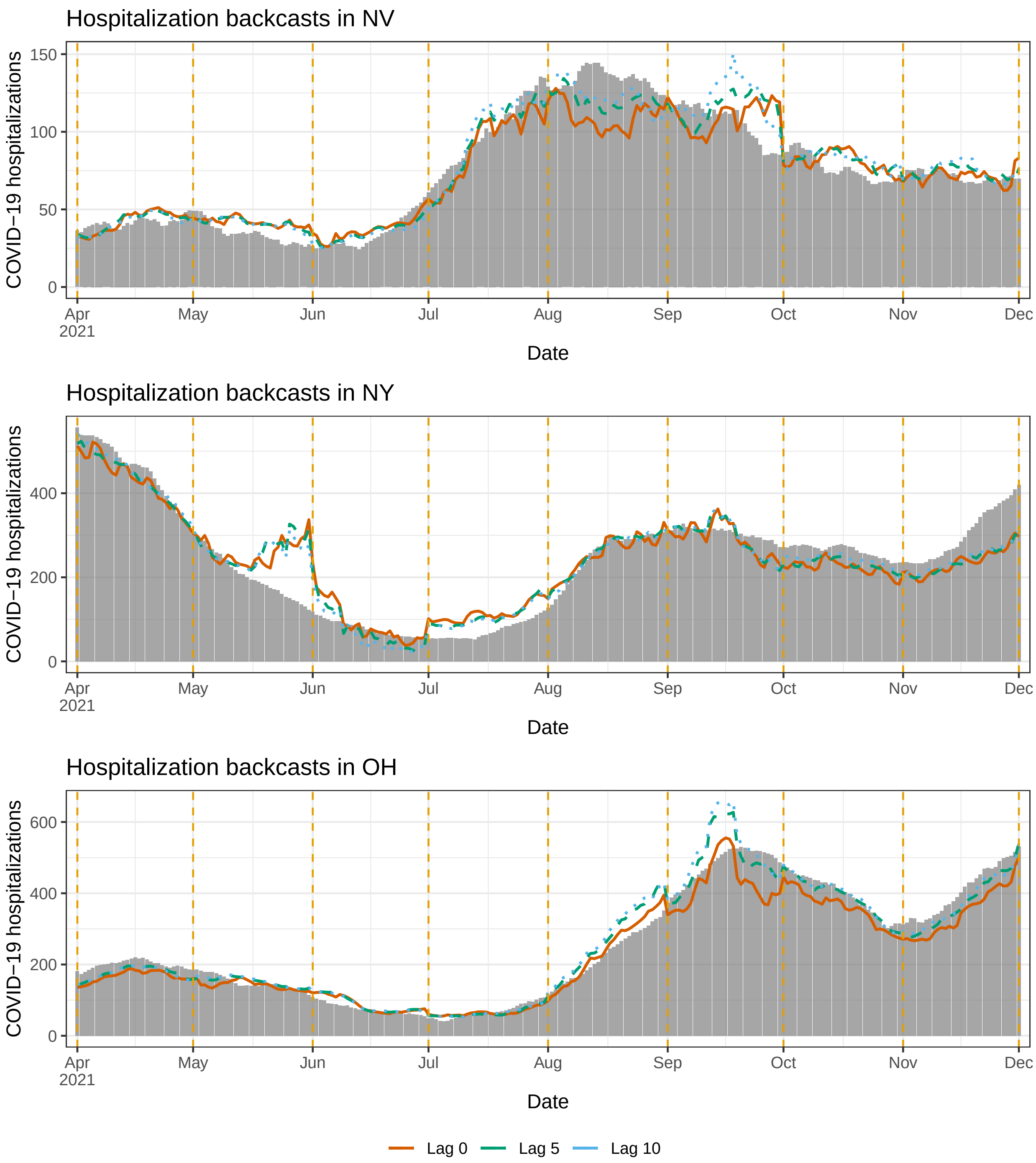

Figure 14: Backcasts from the mixed model in scenario 1, for NV, NY, OH.

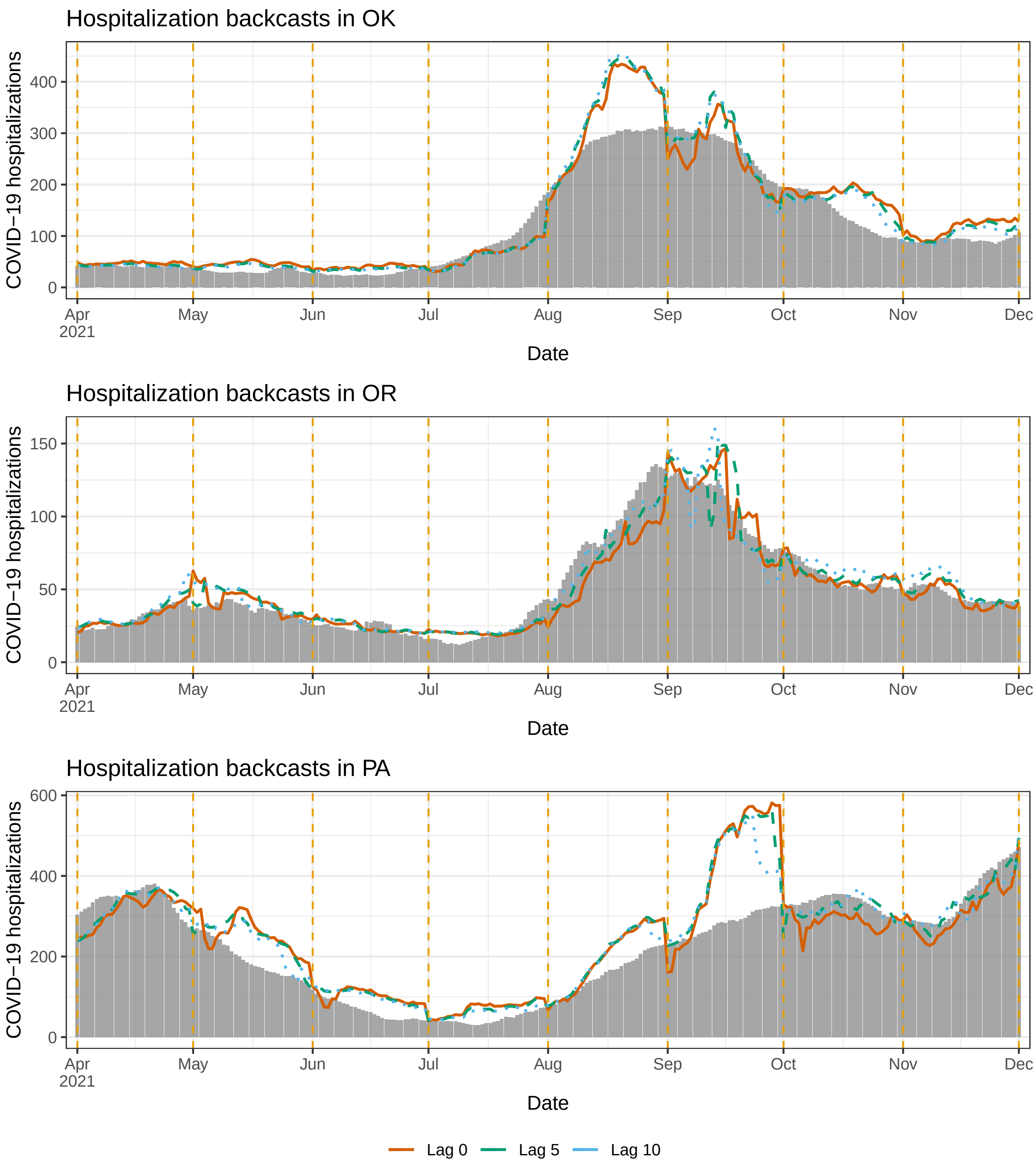

Figure 15: Backcasts from the mixed model in scenario 1, for OK, OR, PA.

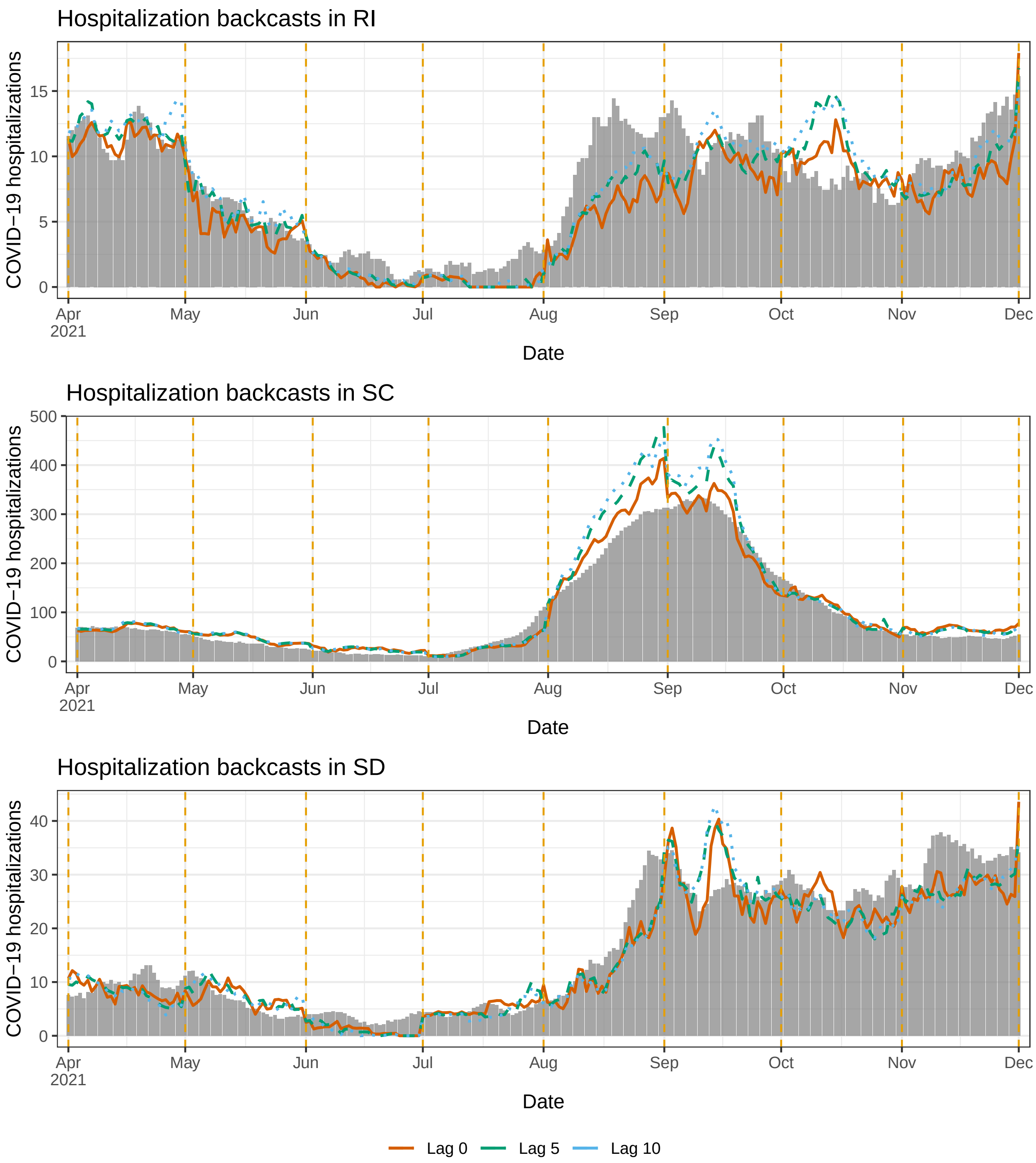

Figure 16: Backcasts from the mixed model in scenario 1, for RI, SC, SD.

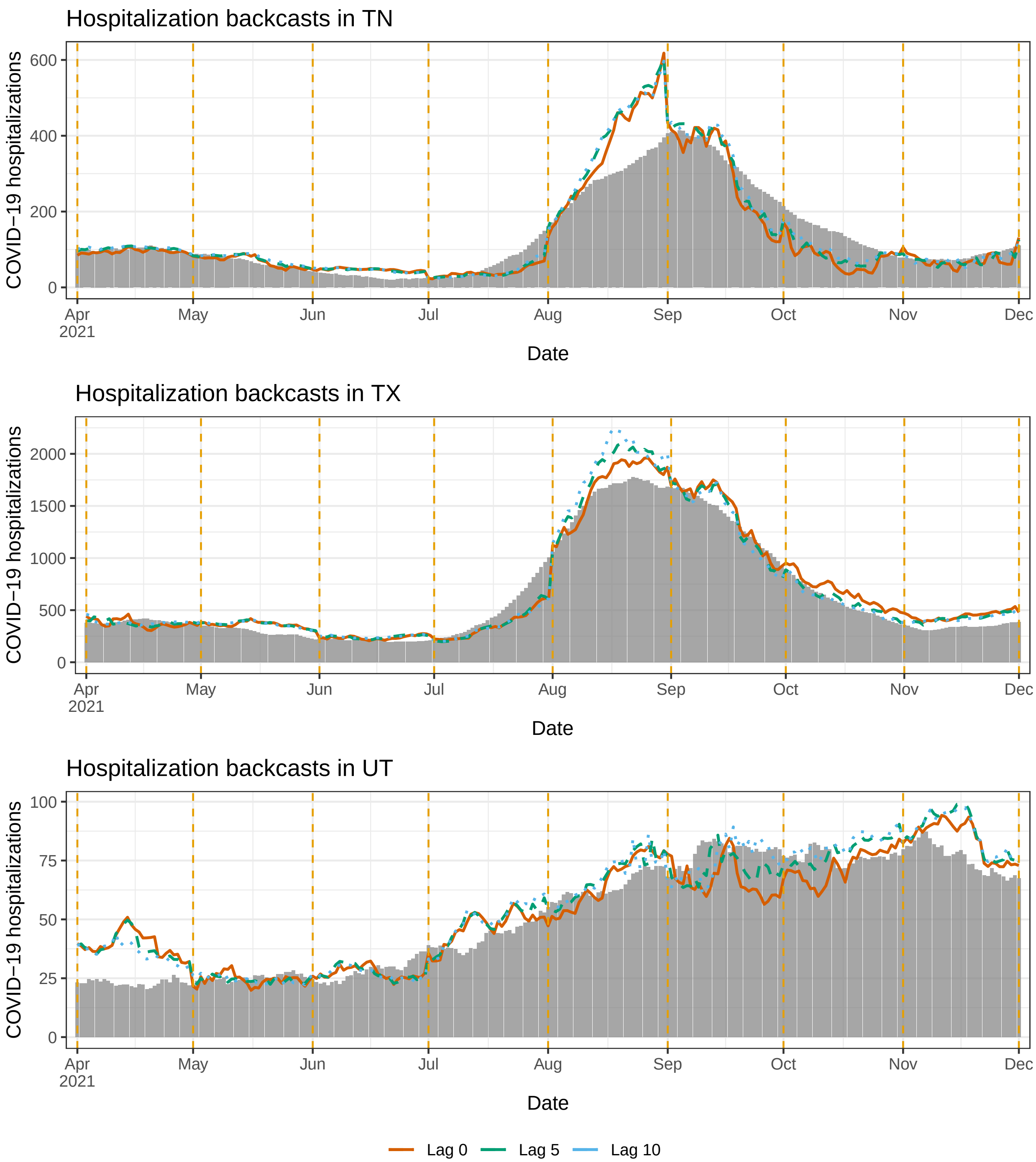

Figure 17: Backcasts from the mixed model in scenario 1, for TN, TX, UT.

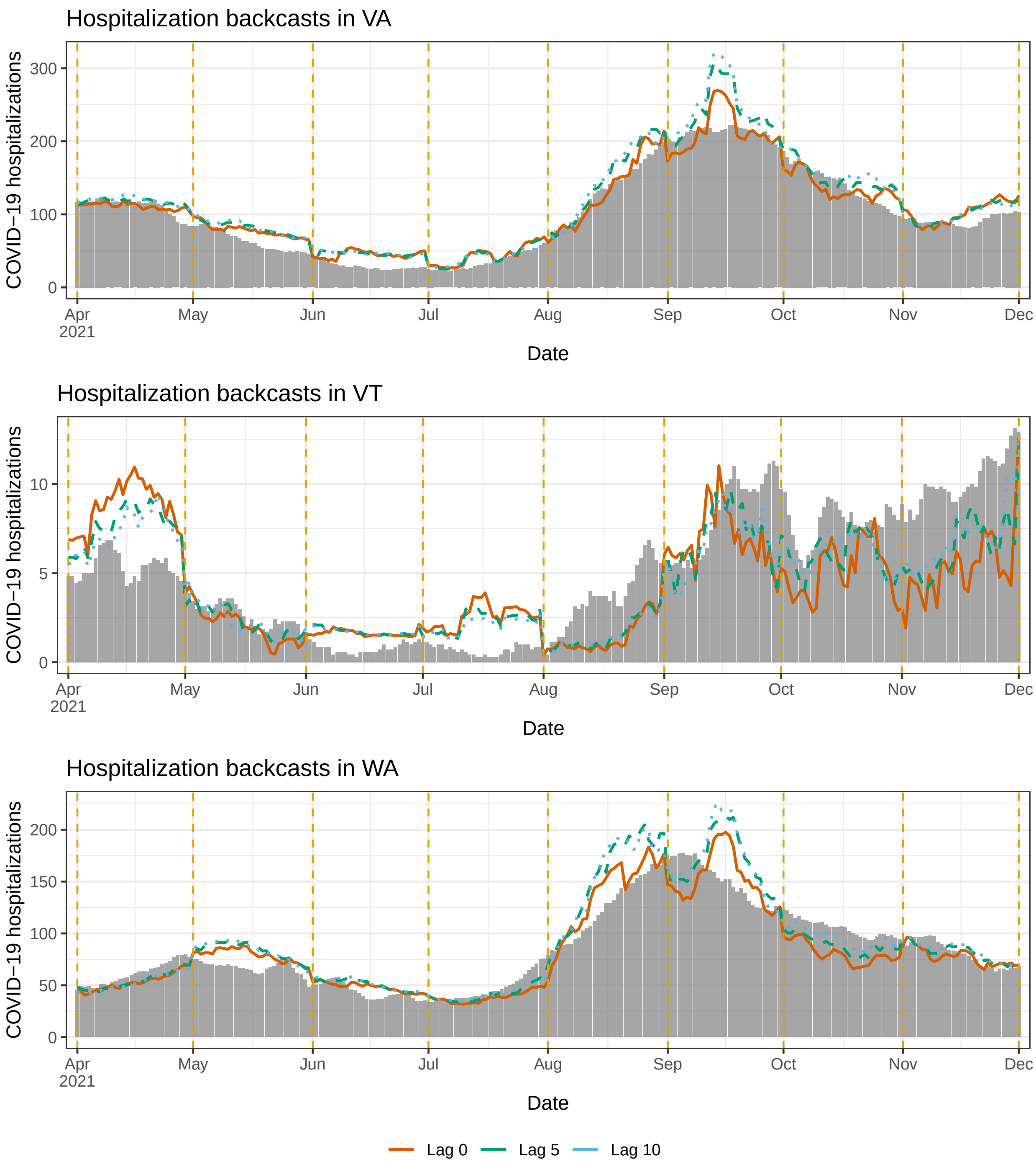

Figure 18: Backcasts from the mixed model in scenario 1, for VA, VT, WA.

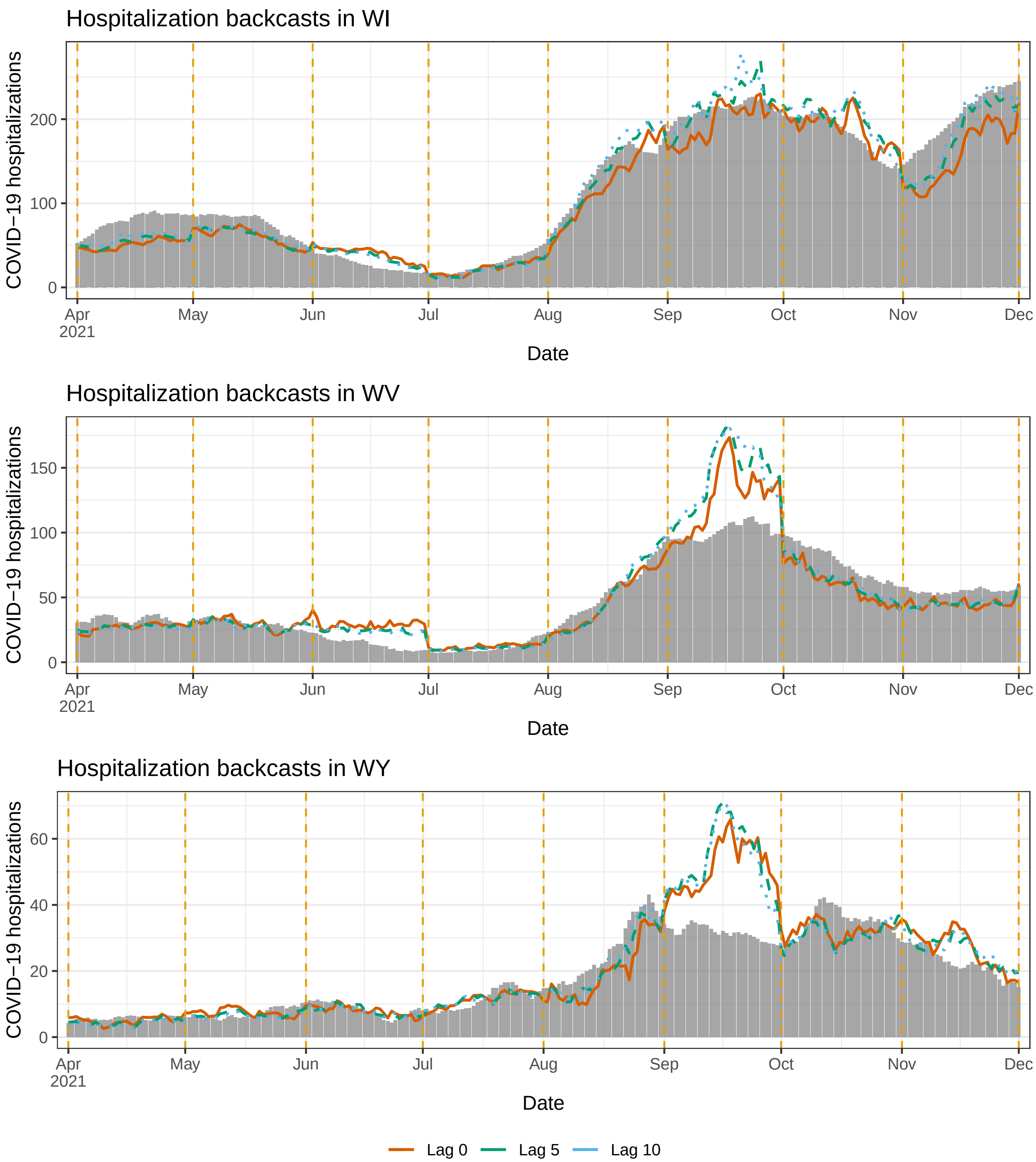

Figure 19: Backcasts from the mixed model in scenario 1, for WI, WV, WY.

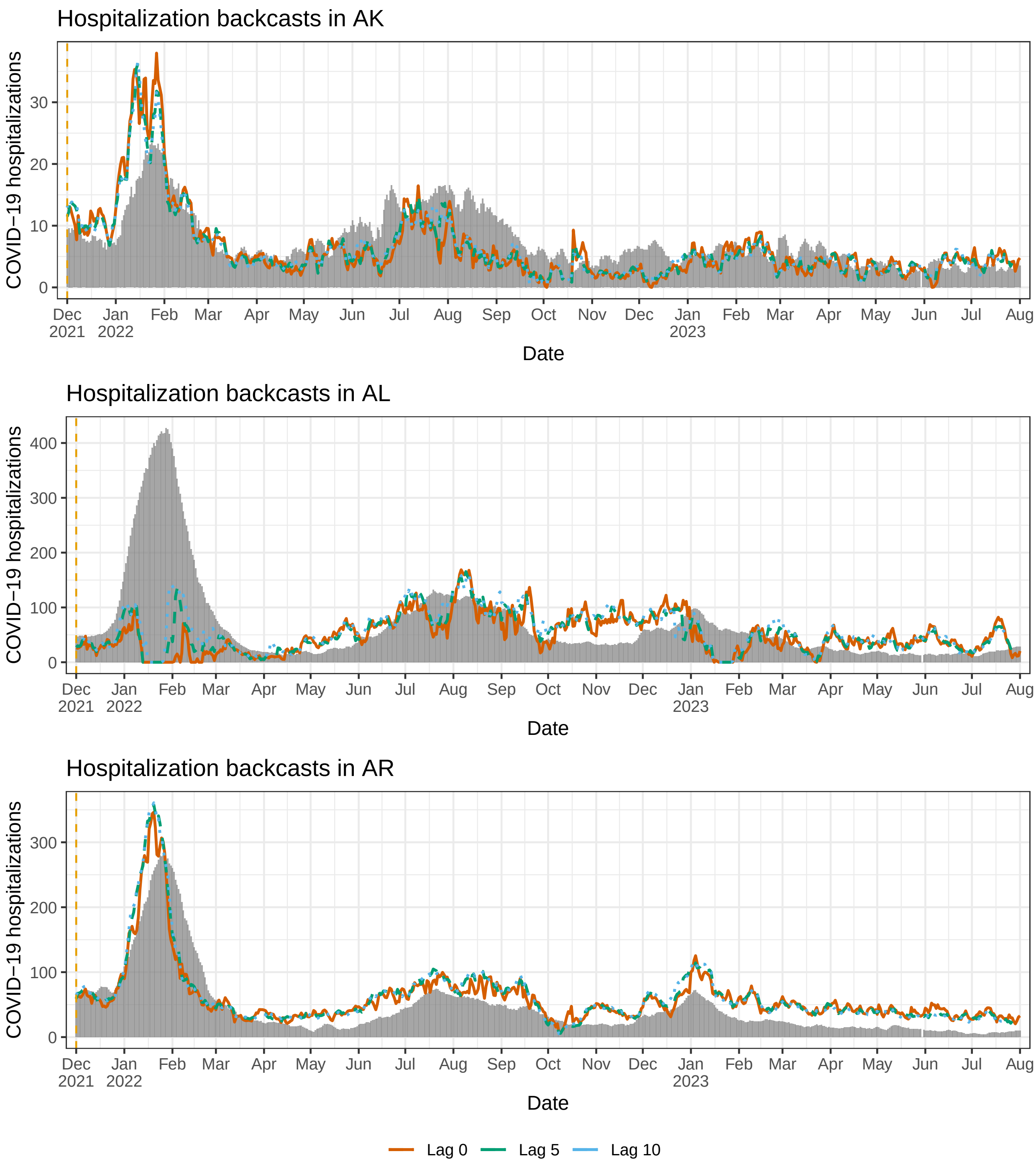

Figure 20: Backcasts from the mixed model in scenario 2, for AL, AK, AZ.

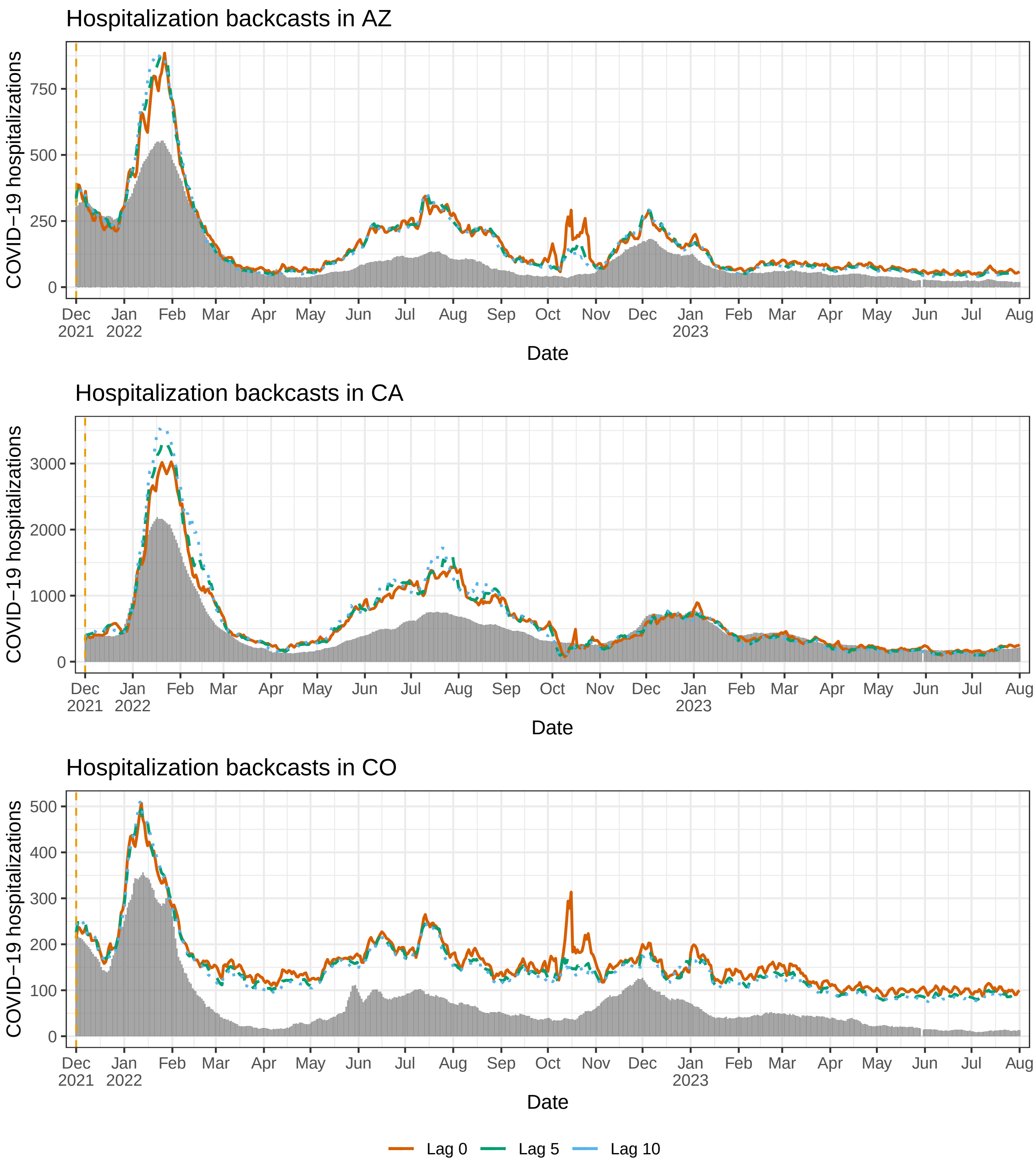

Figure 21: Backcasts from the mixed model in scenario 2, for AR, CA, CO.

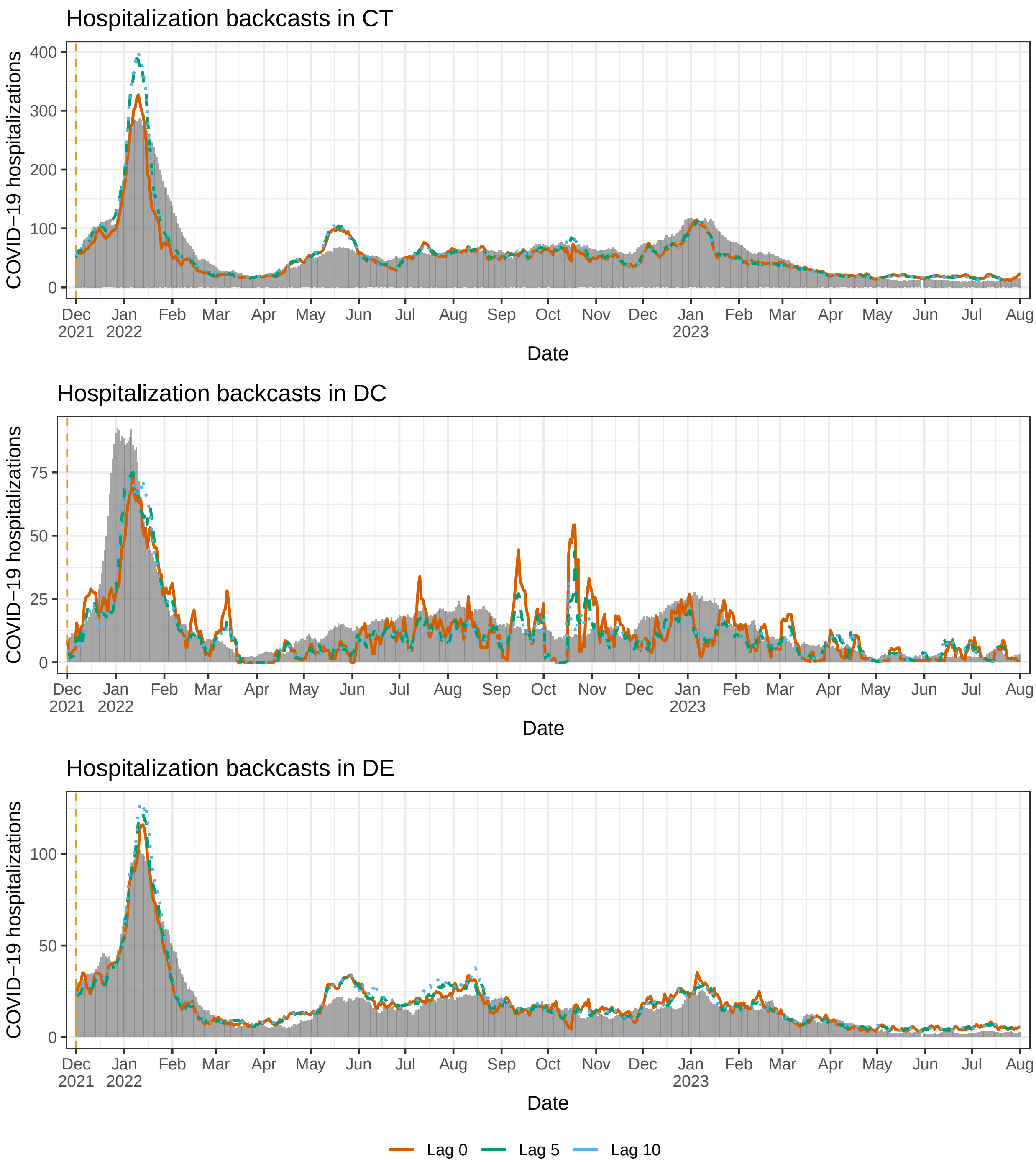

Figure 22: Backcasts from the mixed model in scenario 2, for CT, DC, DE.

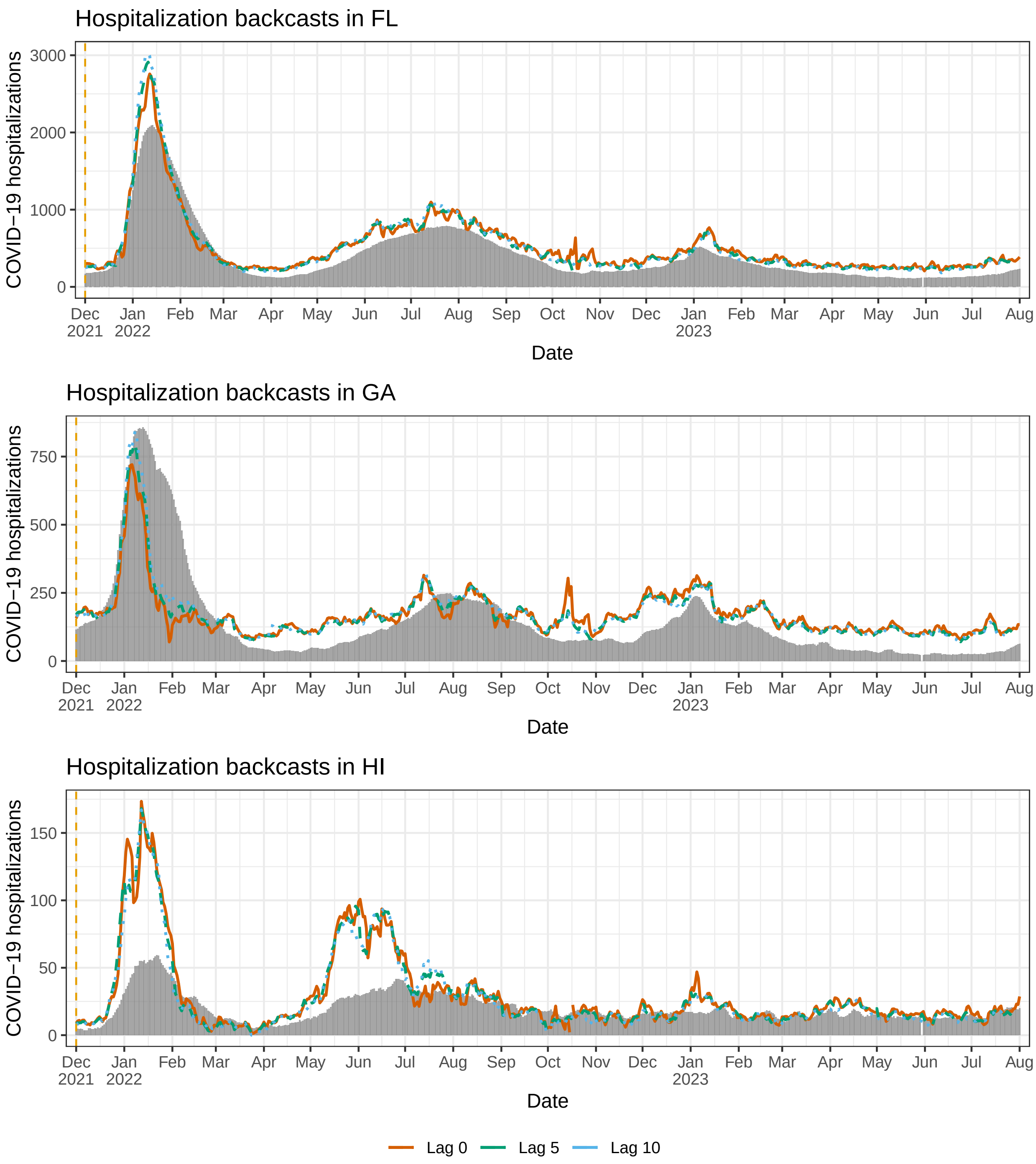

Figure 23: Backcasts from the mixed model in scenario 2, for FL, GA, HI.

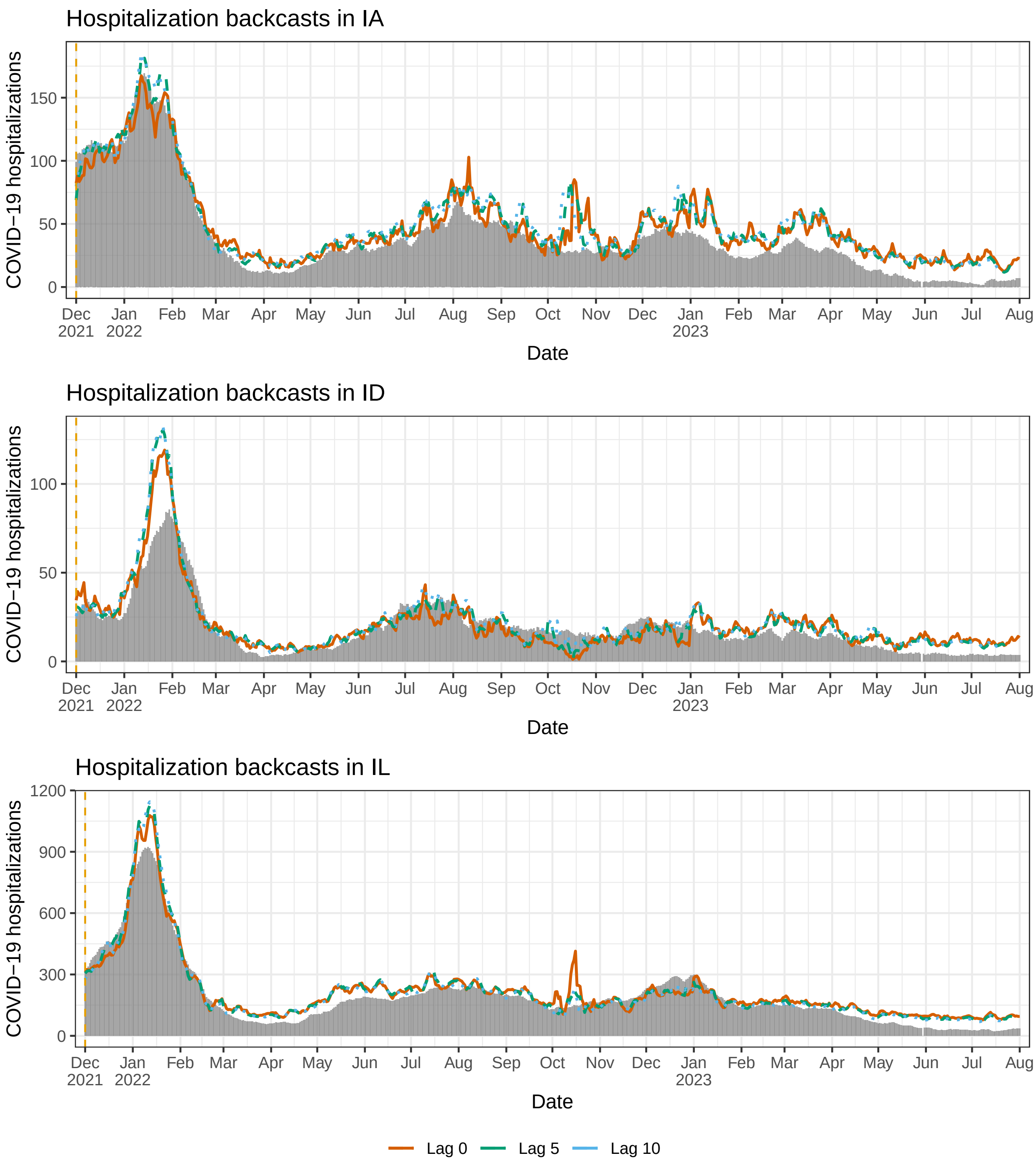

Figure 24: Backcasts from the mixed model in scenario 2, for IA, ID, IL.

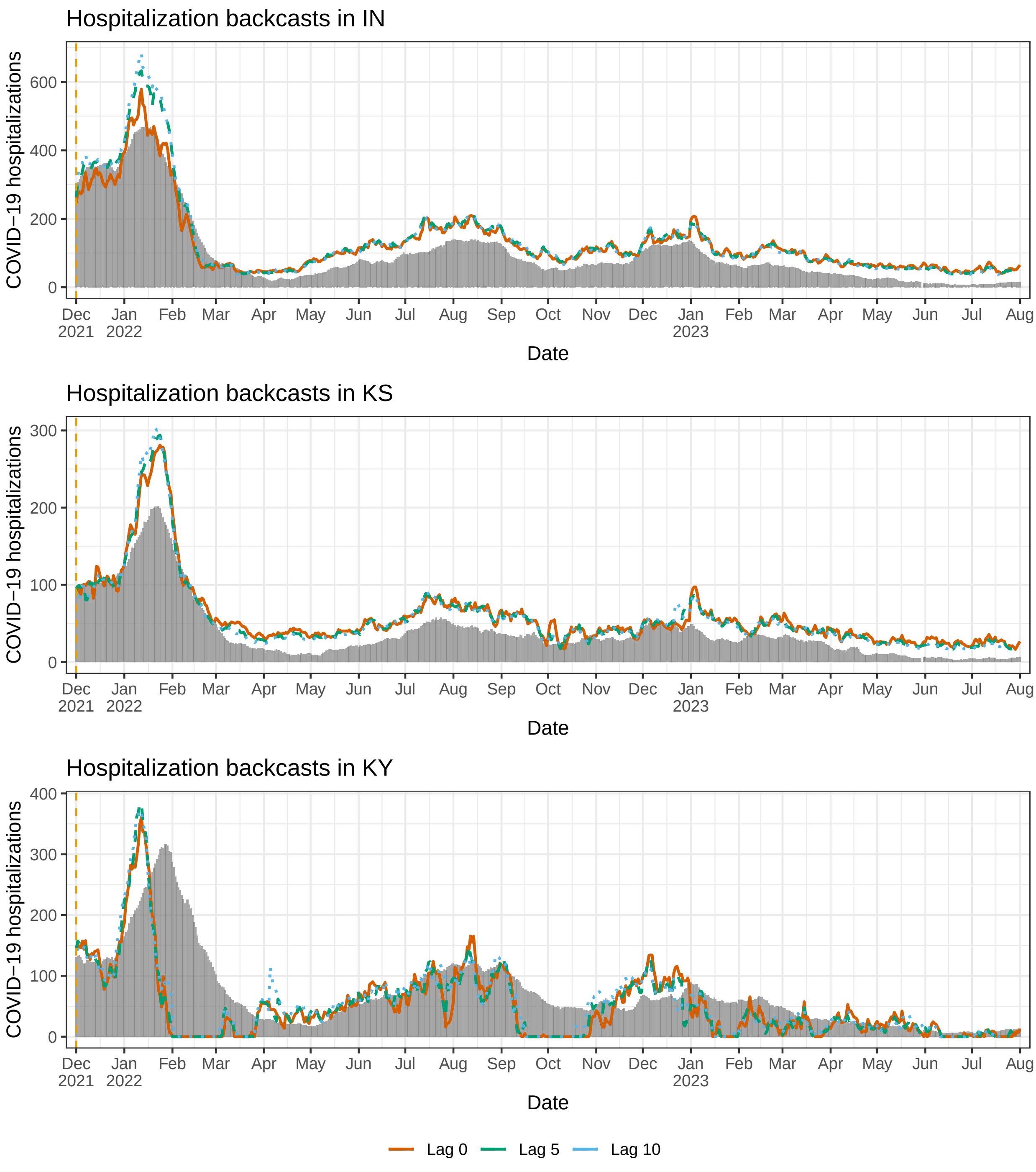

Figure 25: Backcasts from the mixed model in scenario 2, for IN, KS, KY.

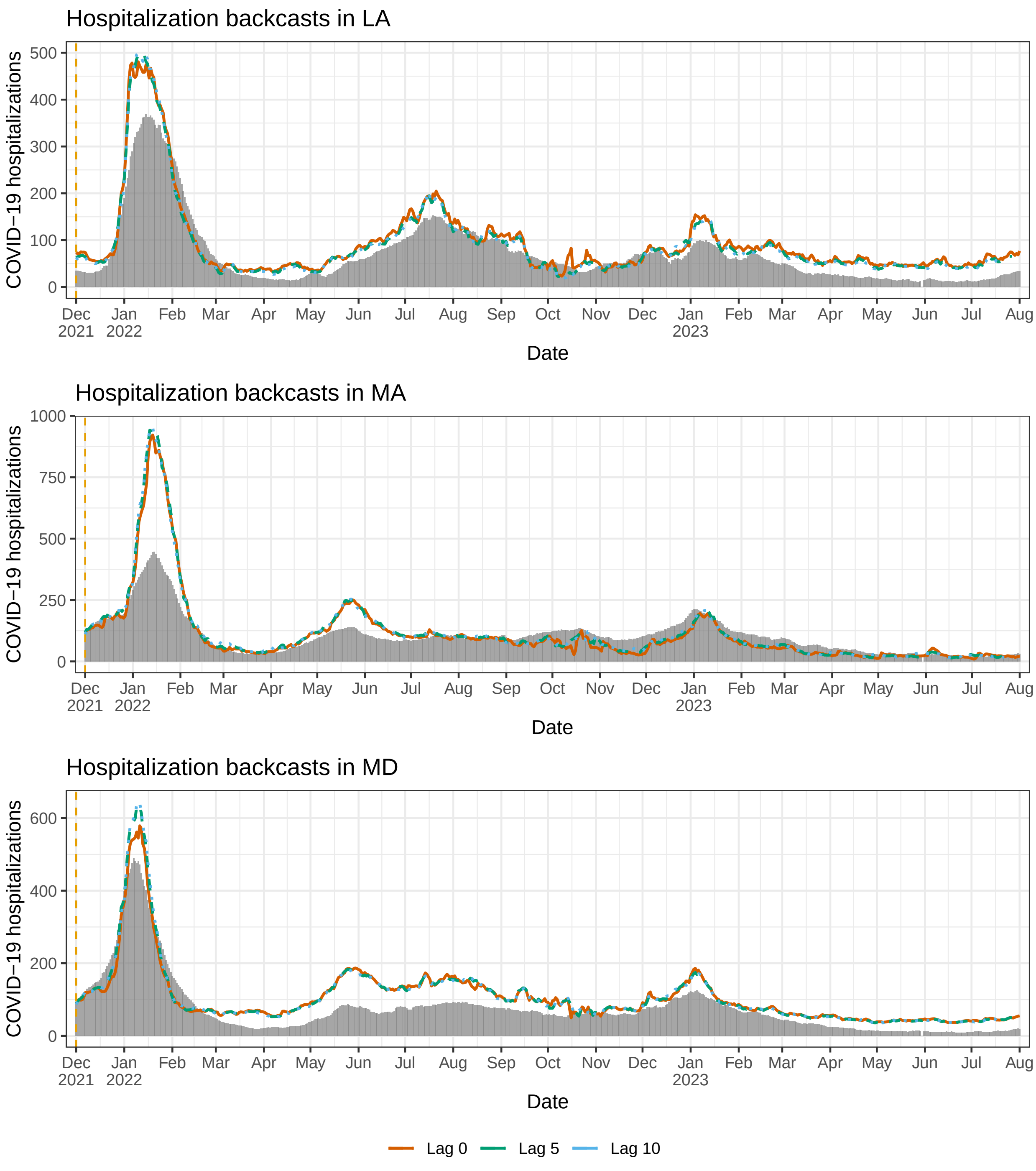

Figure 26: Backcasts from the mixed model in scenario 2, for LA, MA, MD.

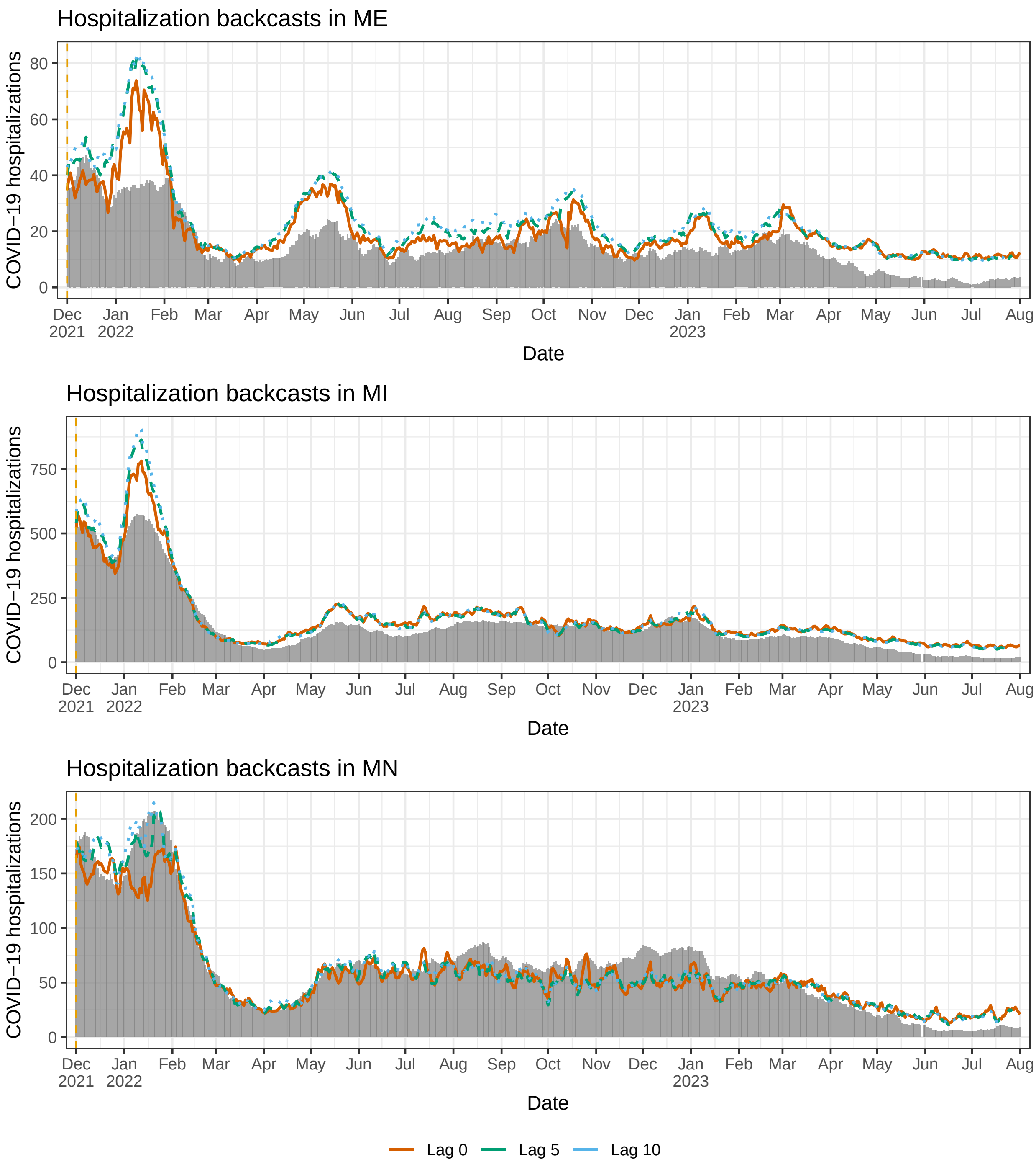

Figure 27: Backcasts from the mixed model in scenario 2, for ME, MI, MN.

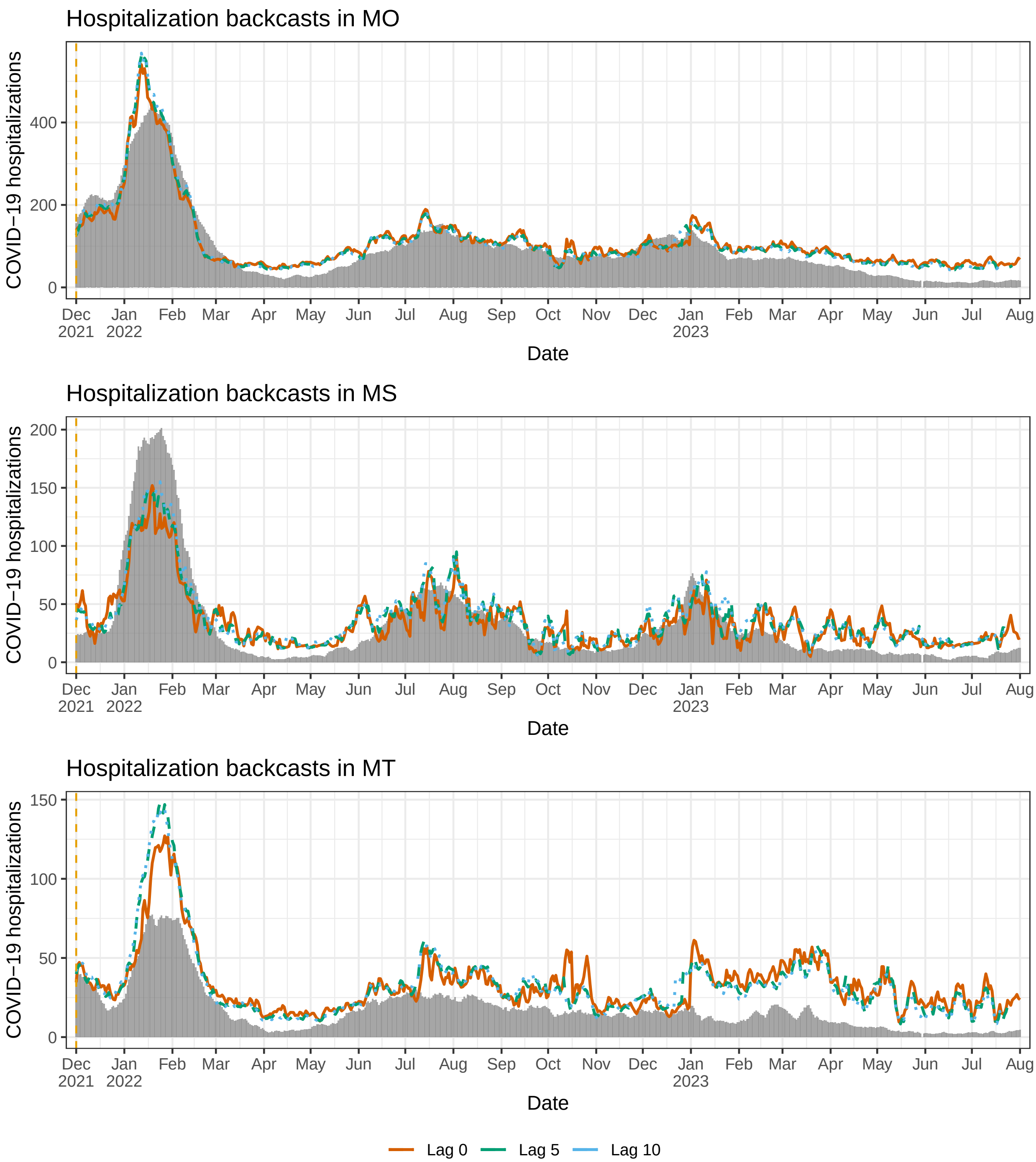

Figure 28: Backcasts from the mixed model in scenario 2, for MO, MS, MT.

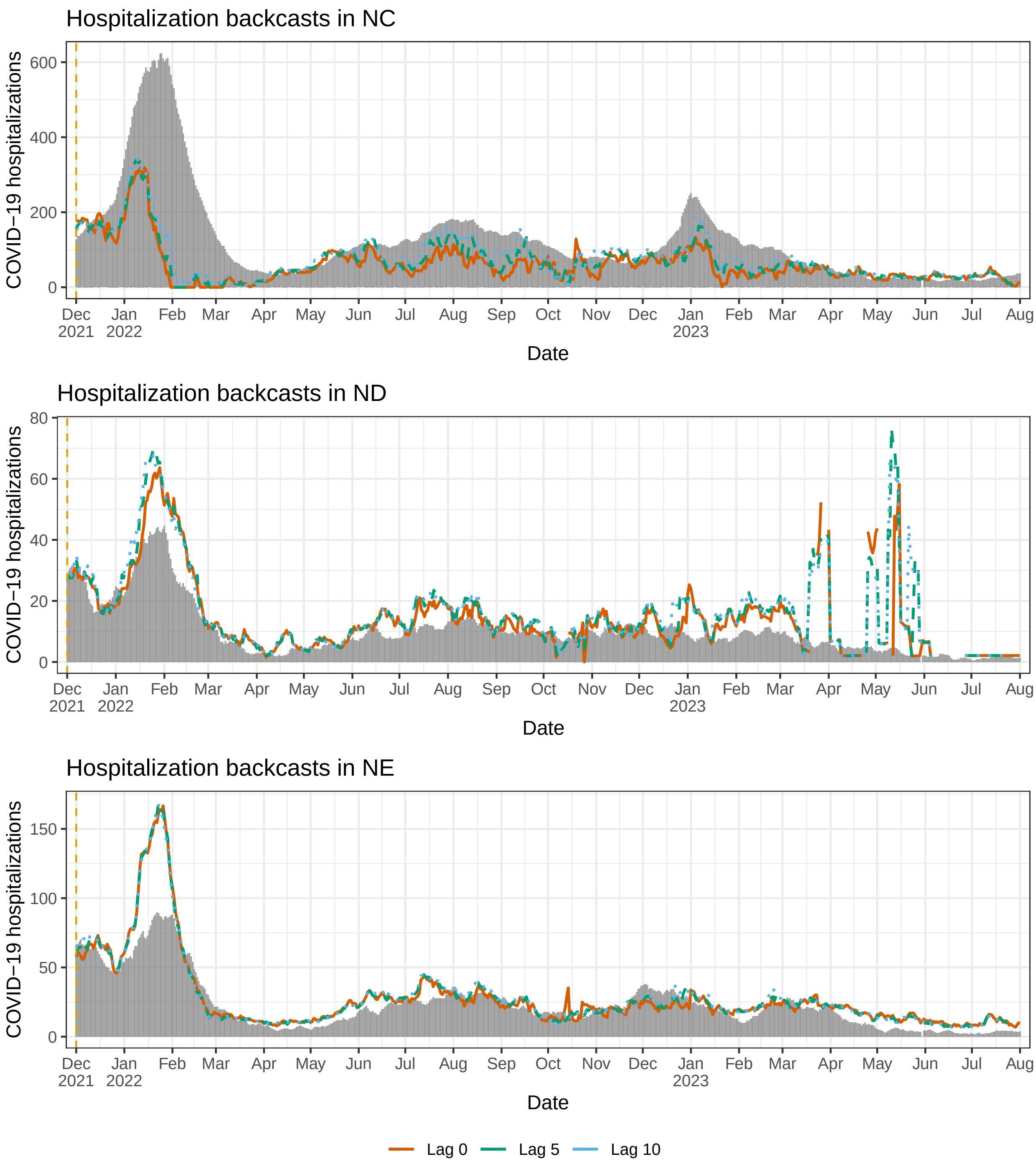

Figure 29: Backcasts from the mixed model in scenario 2, for NC, ND, NE.

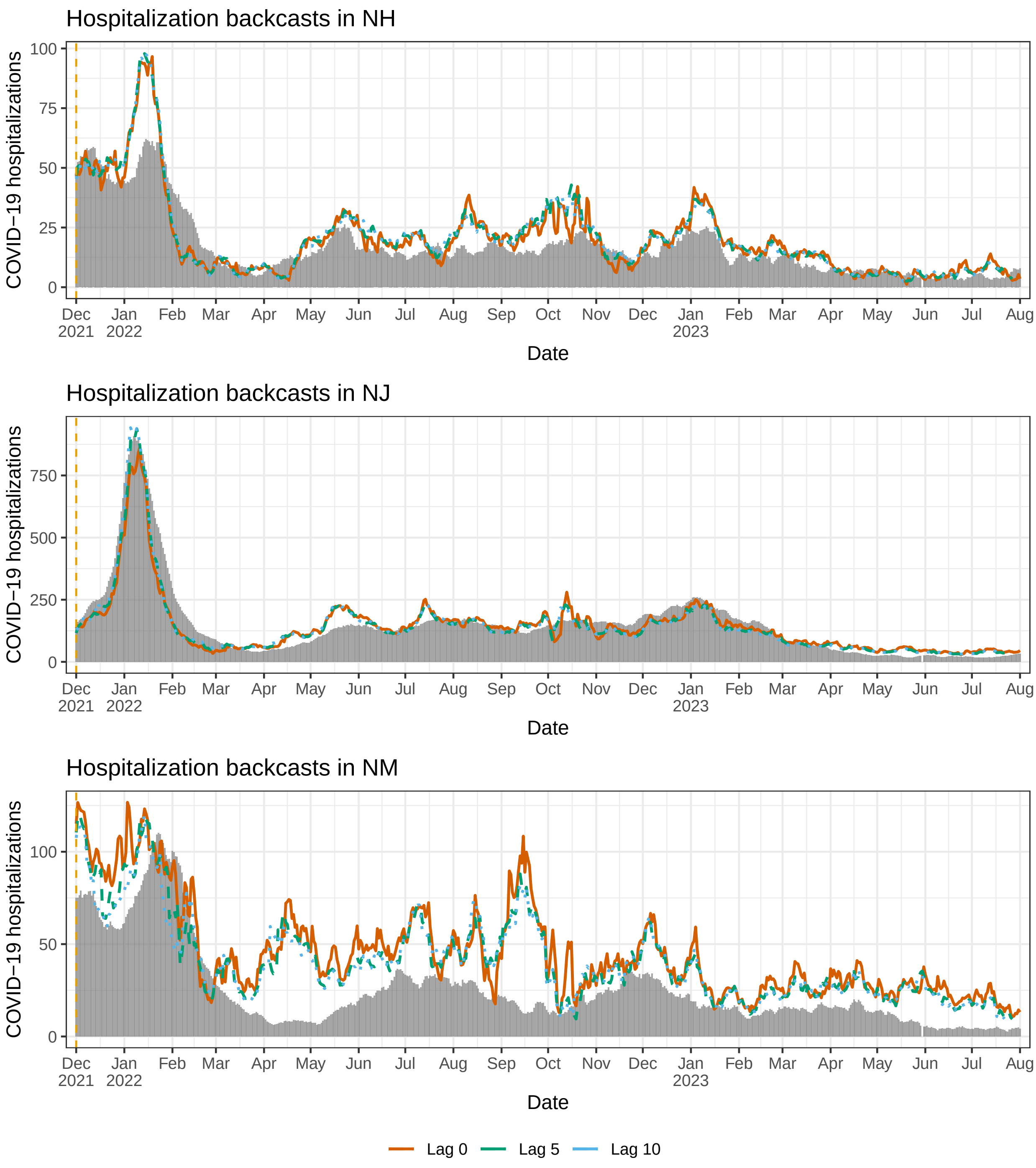

Figure 30: Backcasts from the mixed model in scenario 2, for NH, NJ, NM.

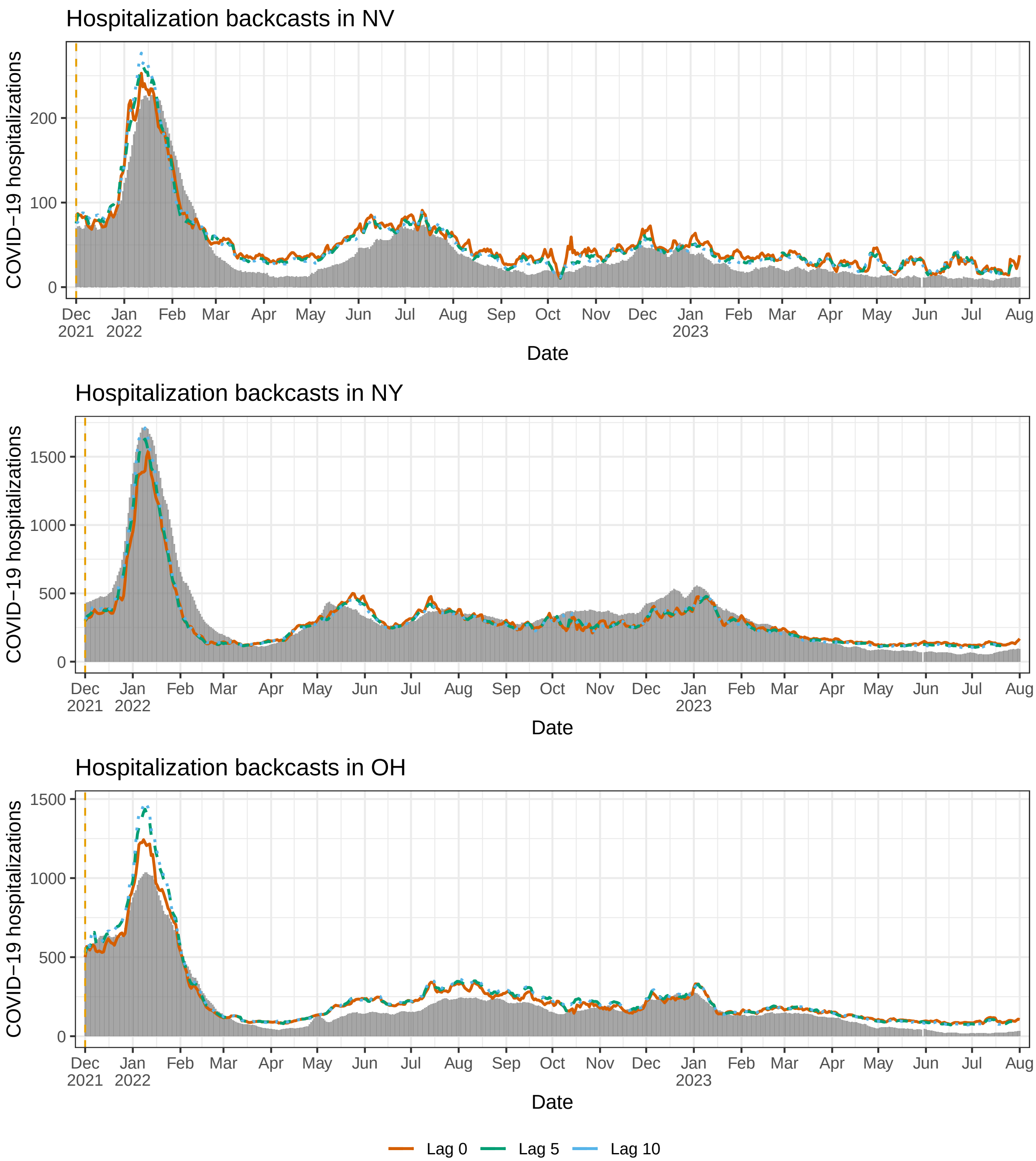

Figure 31: Backcasts from the mixed model in scenario 2, for NV, NY, OH.

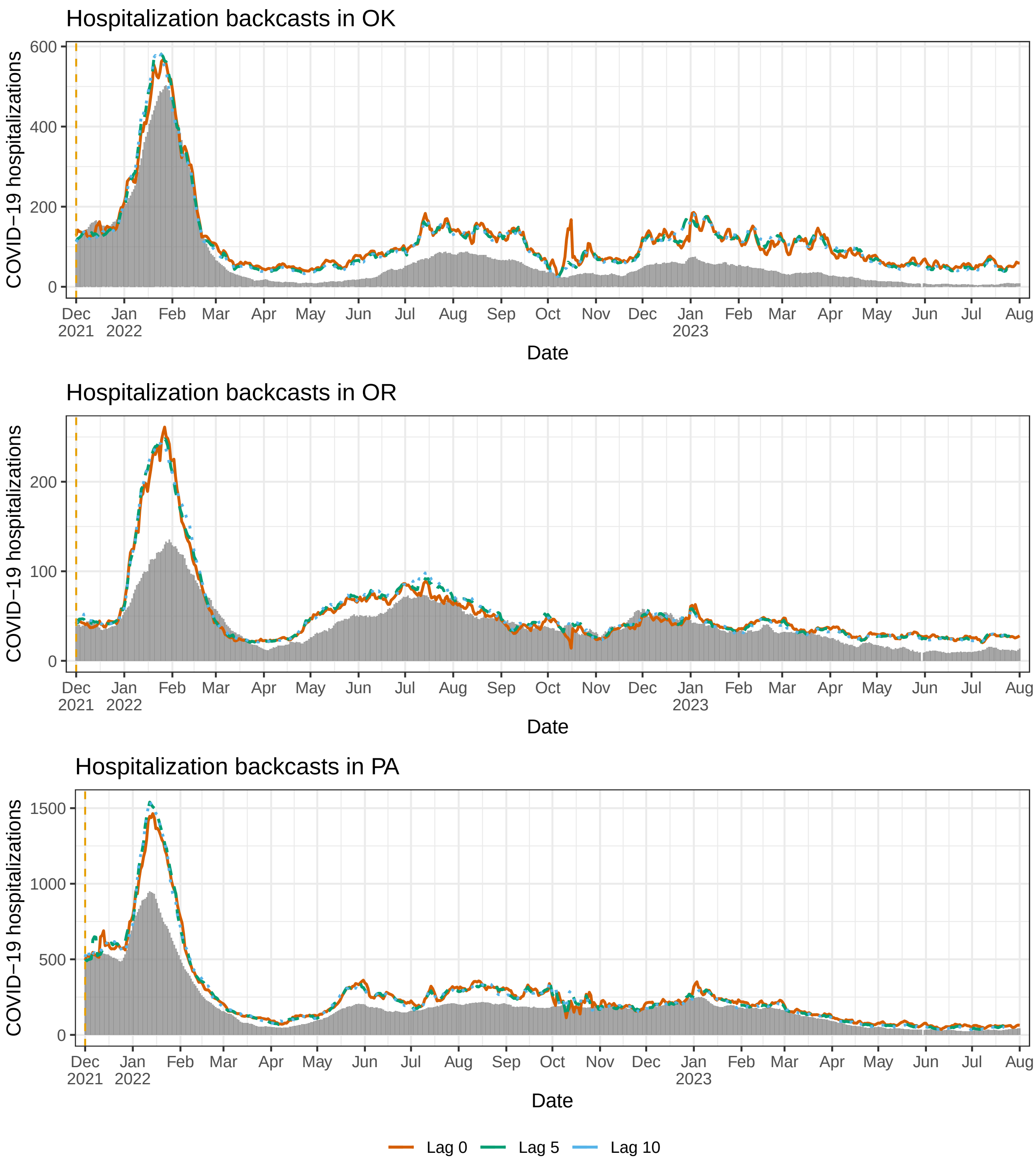

Figure 32: Backcasts from the mixed model in scenario 2, for OK, OR, PA.

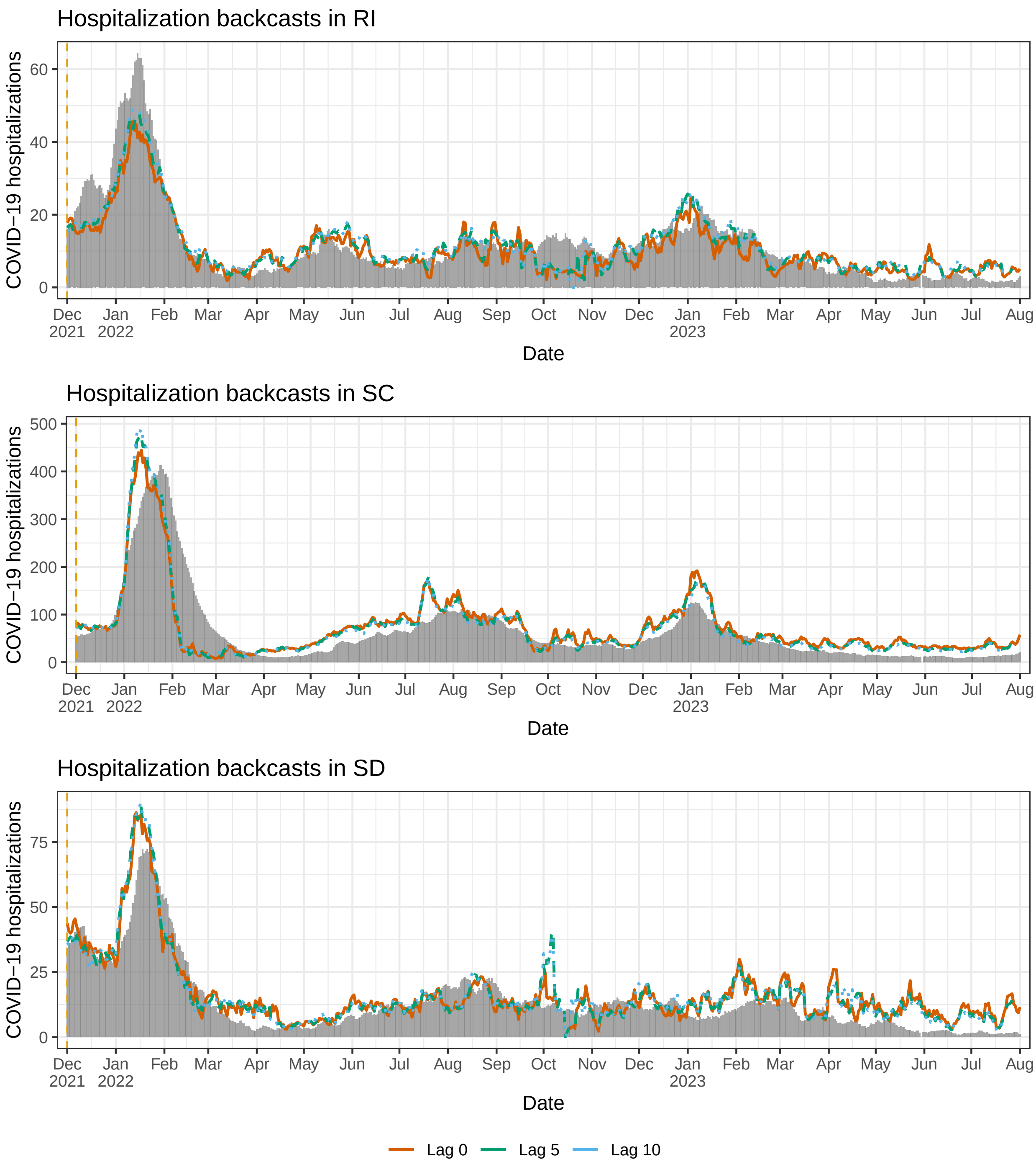

Figure 33: Backcasts from the mixed model in scenario 2, for RI, SC, SD.

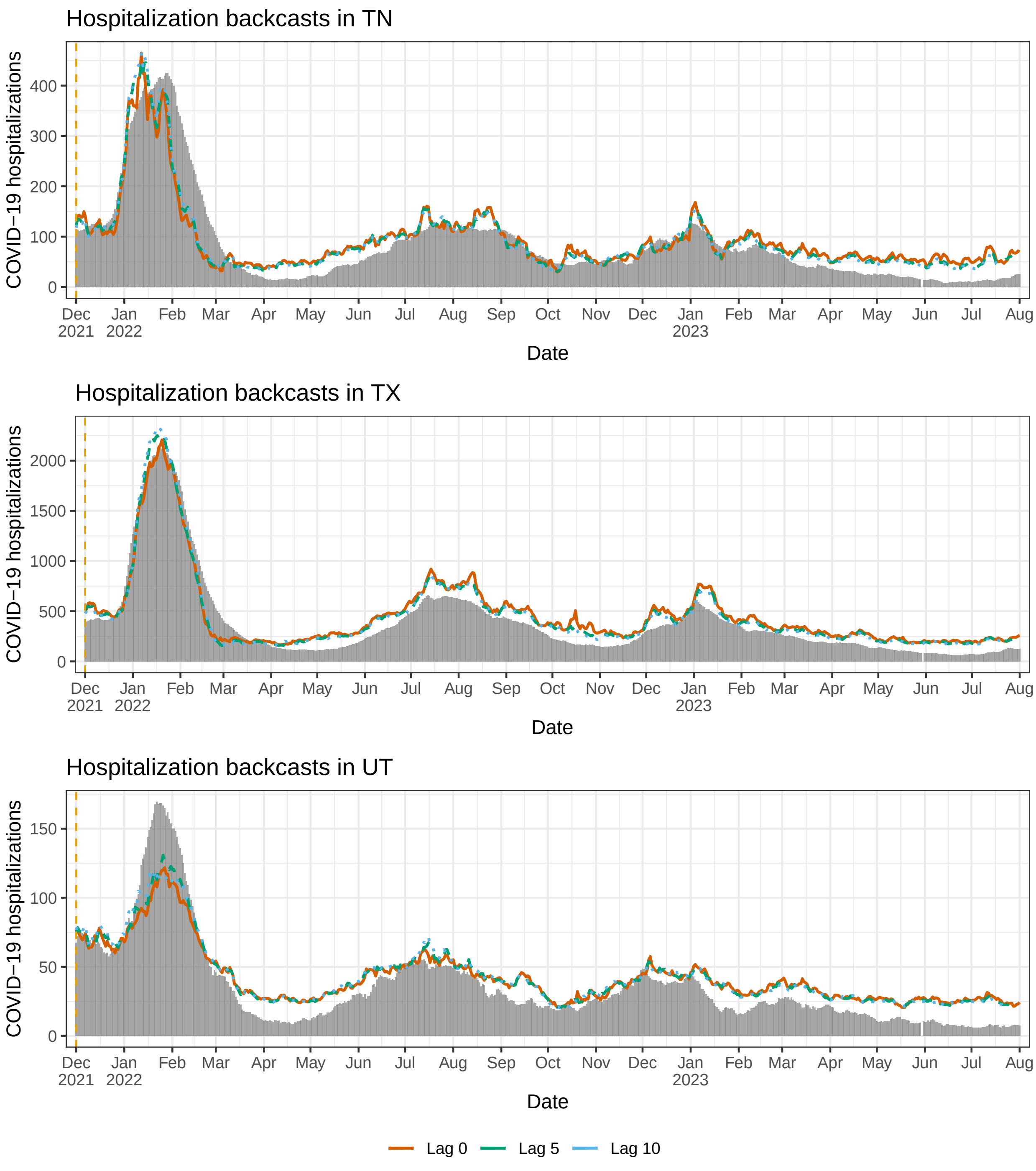

Figure 34: Backcasts from the mixed model in scenario 2, for TN, TX, UT.

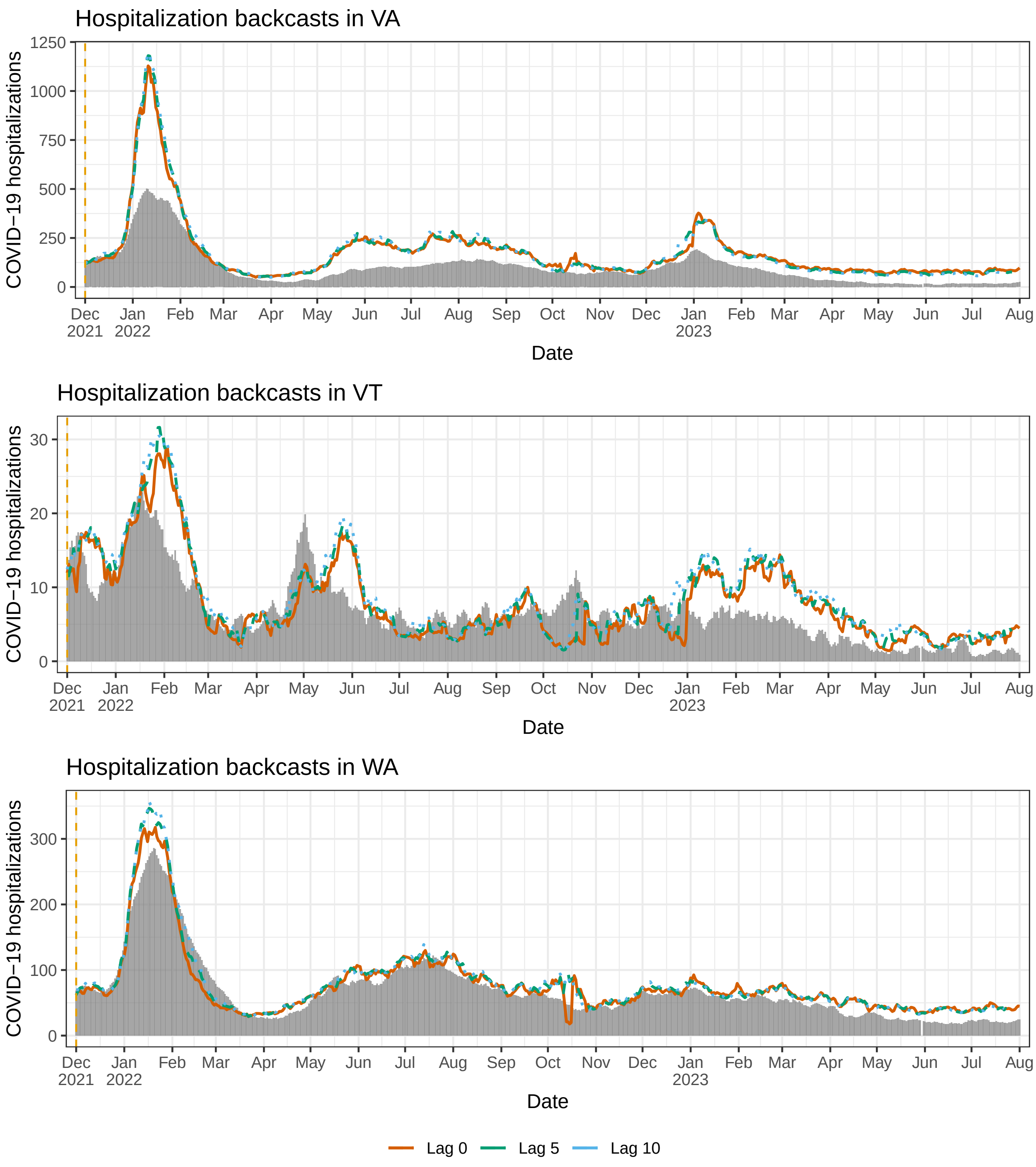

Figure 35: Backcasts from the mixed model in scenario 2, for VA, VT, WA.

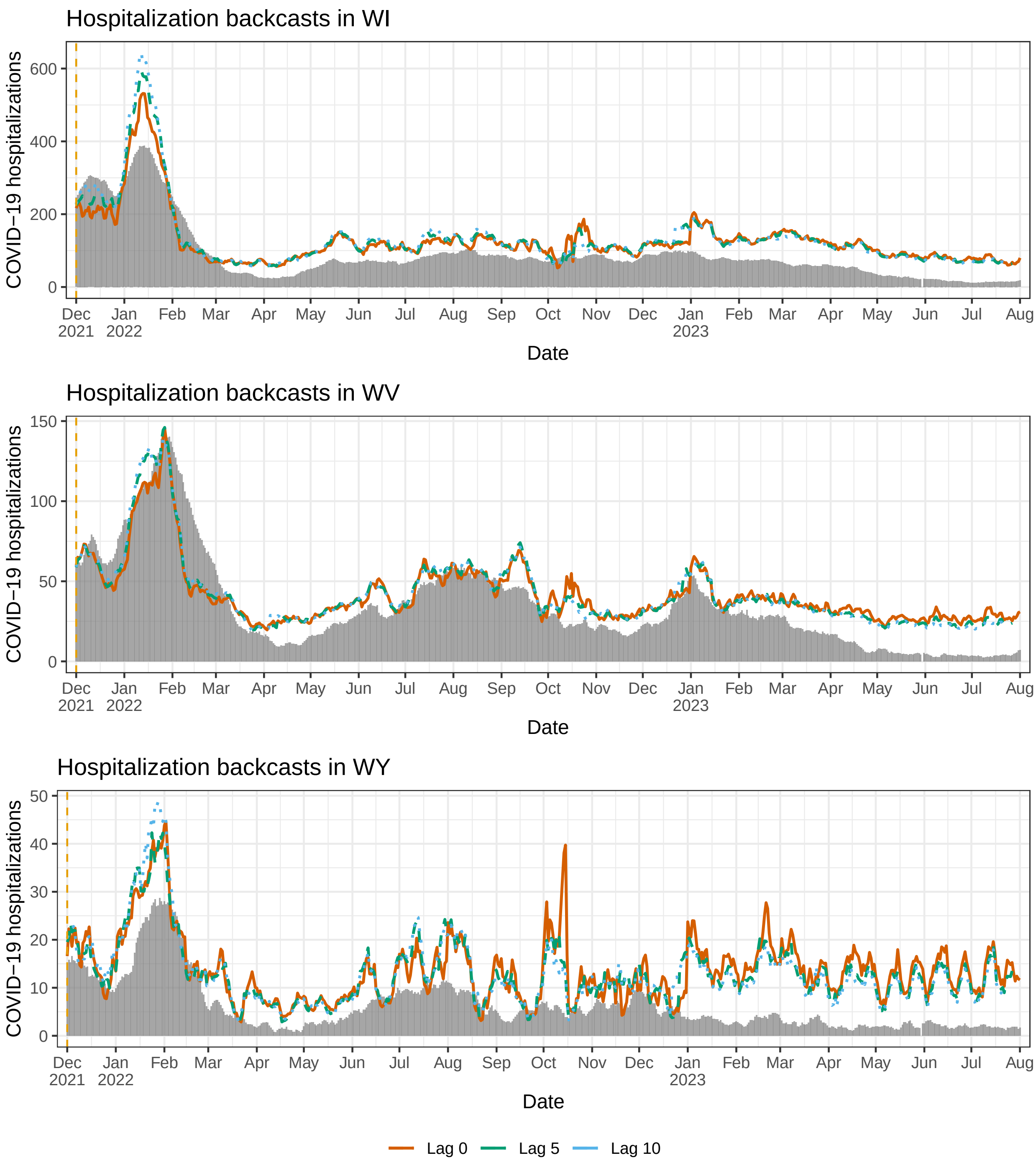

Figure 36: Backcasts from the mixed model in scenario 2, for WI, WV, WY.