Parameter-Free Tour of the Binary Black Hole Population

Thomas A. Callister[1,2,*] and Will M. Farr[2,3]

[1]*Kavli Institute for Cosmological Physics, University of Chicago,*
5640 S. Ellis Avenue, Chicago, Illinois 60615, USA
[2]*Center for Computational Astrophysics, Flatiron Institute,*
162 Fifth Avenue, New York, New York 10010, USA
[3]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA*

The continued operation of the Advanced LIGO and Advanced Virgo gravitational-wave detectors is enabling the first detailed measurements of the mass, spin, and redshift distributions of the merging binary black hole population. Our present knowledge of these distributions, however, is based largely on strongly parametric models. Such models typically assume the distributions of binary parameters to be superpositions of "building block" features like power laws, peaks, dips, and breaks. Although this approach has yielded great progress in the initial characterization of the compact binary population, the strong assumptions entailed often leave it unclear which physical conclusions are driven by observation and which by the specific choice of model. In this paper, we instead model the merger rate of binary black holes as an unknown autoregressive process over the space of binary parameters, allowing us to measure the distributions of binary black hole masses, redshifts, component spins, and effective spins with near-complete agnosticism. We find the primary mass spectrum of binary black holes to be doubly peaked, with a fairly flat continuum that steepens at high masses. We identify signs of unexpected structure in the redshift distribution of binary black holes: a uniform-in-comoving volume merger rate at low redshift followed by an increase in the merger rate beyond redshift $z \approx 0.5$. Finally, we find that the distribution of black hole spin magnitudes is unimodal and concentrated at small but nonzero values, and that spin orientations span a wide range of spin-orbit misalignment angles but are also moderately unlikely to be truly isotropic.

I. BACKGROUND

The recent release of the third gravitational-wave transient catalog (GWTC-3) [1] by the LIGO Scientific Collaboration [2], Virgo Collaboration [3], and KAGRA Collaboration [4] has increased the number of confident gravitational-wave detections to 76 [5], with yet more candidates identified in independent reanalyses of LIGO-Virgo data [7,8]. This growing body of detections has pushed gravitational-wave astronomy firmly into the catalog era; we can move beyond interrogating the properties of individual binary mergers to instead exploring the ensemble properties of the complete compact binary population [6,9].

Most present-day analyses of the compact binary population adopt a strongly modeled approach, in which the distributions of binary masses, spins, and redshifts are assumed to follow specific parametric forms. The binary black hole mass spectrum, for example, is commonly assumed to be the superposition of power laws and/or Gaussians [6,9–14]. The hyperparameters describing these functional forms (e.g., power-law slopes and Gaussian means and widths) are then measured using our catalog of gravitational-wave detections.

This strongly modeled "building-block" approach has yielded significant insight. We have learned, for example, that the black hole merger rate is highest at $m \approx 10 M_\odot$, declining steeply towards larger masses but with a secondary bump near $35 M_\odot$ [6,9]. Black hole spins are small but nonzero, with a wide range of misalignment angles between spins and binary orbital angular momenta [6,9,15,16]. In addition, the rate of binary mergers grows as we look to higher redshifts [6].

At the same time, this approach has some less-desirable downsides.

(i) First, our chosen functional form prescribes, from the very outset, the set of possible population features. Thus, it is not always clear which conclusions come

*[*]tcallister@uchicago.edu

from informative data and which are built, by assumption, into the models themselves. Parametrized models including sharp features, for example, are prone to "false alarms," favoring the existence of such features even when none exist [15].

(ii) Second, different models may yield very different or even conflicting conclusions if they prescribe different sets of features, further obscuring which conclusions are robust and which are model induced.

(iii) Finally, strongly parametrized models allow us to search for "known unknowns" (e.g., is there a pair-instability cutoff in the black hole mass spectrum?) but do not let us search for the "unknown unknowns," truly unexpected features that might challenge our astrophysical understanding of compact binary formation and evolution. Several features in the binary black hole population (a peak in the merger rate near $35 M_\odot$ [9], a correlation between binary mass ratio and spin [17,18], etc.), for example, were discovered serendipitously only after a fortuitous choice of model.

These concerns have spurred the development of flexible methods that aim to characterize the compact binary population while imposing few *a priori* assumptions regarding the form of the population. Examples of these flexible approaches include modeling the distribution of binary parameters using splines [19–21], piecewise-constant "binned" models [6,22–26], and Gaussian mixture models [13,27,28], as well as non-Bayesian methods that seek to identify clustering in gravitational-wave catalogs [29,30].

In this paper, we will explore an alternative and complementary approach, treating the merger rate of binary black holes as an unknown autoregressive process defined over masses, spins, and/or redshifts. Whereas all other population models entail the use of hyperparameters to specify the dependence of the merger rate on mass, spin, and redshift, under our approach the merger rates at every posterior sample are themselves the quantities that we directly infer from data. This process allows us to characterize the compact binary distribution with a high degree of agnosticism, assuming only a prior preference that the merger rate be a continuous function of binary parameters. This approach will allow us to confirm the robustness of features previously identified using standard strongly parametrized models, as well as identify new features that might otherwise be overlooked.

More specifically, our goals in this work are threefold:

(1) First, we present a flexible measurement of the merger rate densities and probability distributions over binary black hole masses, mass ratios, redshifts, and spins.

(2) Inference of the binary merger rate is only a first step. As we discuss further in Sec. VII, a conceptually distinct and equally important step is feature extraction: the subsequent identification of

features and assessment of their significance. For strongly parametrized models, rate inference and feature extraction are, by definition, performed simultaneously. This is not the case for flexible approaches like spline methods or our autoregressive process, and so the importance of developing techniques for feature extraction is particularly acute. For each binary black hole parameter, we therefore seek to methodically assess feature significance without resorting to hyperparameters, but instead by calculating and comparing merger rates between different regions of interest.

(3) Finally, our goal is to leverage our autoregressive inference to provide new or extended strongly parametrized models that reflect our most up-to-date understanding of the binary black hole population. This goal serves two purposes. First, these updated models may provide a robust and accessible point of comparison between theory and observation. Second, these strongly parametrized models can, in turn, be adopted in traditional hierarchical analyses, helping to confirm (or reject) possible features that appear in more flexible analyses of the black hole population.

In Sec. II, we begin by motivating and defining autoregressive processes as a useful tool in the inference of compact binary populations. In Secs. III–VI, we then apply our method to study the distributions of masses, redshifts, component spins, and effective spins among the binary black hole population. Along the way, we systematically demonstrate feature extraction and discuss new or expanded strongly parametrized models motivated by our results. We conclude in Sec. VII by commenting on the relationship between flexible and strongly parametrized models and avenues for future work.

II. THE COMPACT BINARY POPULATION AS AN AUTOREGRESSIVE PROCESS

A. Autoregressive processes

To help make our discussion concrete, consider the problem of measuring the binary black hole primary mass distribution. This amounts to measuring the differential merger rate

$$\frac{dR}{d\ln m} \equiv \frac{dN}{dV_c dt_s d\ln m}, \tag{1}$$

giving the number dN of mergers per unit comoving volume dV_c, per unit source-frame time dt_s, and per logarithmic mass interval $d\ln m$. For notational convenience, we will use the shorthand $\mathcal{R}(\theta) \equiv dR/d\theta$ to denote the merger rate density over parameters θ, e.g.,

$$\mathcal{R}(\ln m) \equiv \frac{dR}{d\ln m}. \tag{2}$$

The standard strongly parametrized approach involves assuming some particular functional form for $\mathcal{R}(\ln m)$, such as a superposition of power laws, Gaussians, and/or truncations, and then measuring the parameters of these functions [6,9–14]. Stepping back, however, we can think more generally about the merger rate density $\mathcal{R}(\ln m)$ that we seek to measure.

In nature, there exists *some* underlying function that describes the true mass spectrum of compact binaries; this function is illustrated in cartoon form by the dark blue curve in Fig. 1. We know nothing *a priori* about the exact shape of this function. However, we can still attempt to write down prior assumptions about this function's likely behavior. In Fig. 1, we hypothetically know the merger rate $\mathcal{R}_i$ at some particular value $\ln m_i$. Given this knowledge, what is our prior expectation on the merger rate at a new point $\ln m_{i+1}$? A reasonable expectation is that, if $\ln m_i$ and $\ln m_{i+1}$ are close together (Scenario 1 in the top panel), then the rates at these locations are likely similar as well. In fact, in the limit that $\ln m_i = \ln m_{i+1}$, we should recover $\mathcal{R}_i = \mathcal{R}_{i+1}$. Conversely, if $\ln m_i$ and $\ln m_{i+1}$ are far apart (Scenario 2), then the rates at each point need not be similar at all.

This intuition forms the basis of an autoregressive process prior. An autoregressive process $\Psi(x)$ is a stochastic function whose value Ψ_i at some new point is related to the values at all previous points by

$$\Psi_i = \sum_{j=1}^{p} c_j \Psi_{i-j} + w_i. \tag{3}$$

Here, the $\{c_j\}$ are deterministic coefficients, and w_i is a random variable. Qualitatively, the coefficients $\{c_j\}$ govern the degree to which $\Psi(x)$ "remembers" its past values, while the parent distribution of $\{w_i\}$ governs the degree to which the function is allowed to randomly fluctuate. The parameter p is called the "order" of the process, and determines the smoothness of the resulting functions; an autoregressive process of order p has $p-1$ continuous derivatives. Choosing order $p = 1$ gives us the simplest "AR(1)" autoregressive process, which obeys

$$\Psi_i = c_i \Psi_{i-1} + w_i. \tag{4}$$

We can adopt this language as a framework with which to codify our intuition regarding possible merger rate
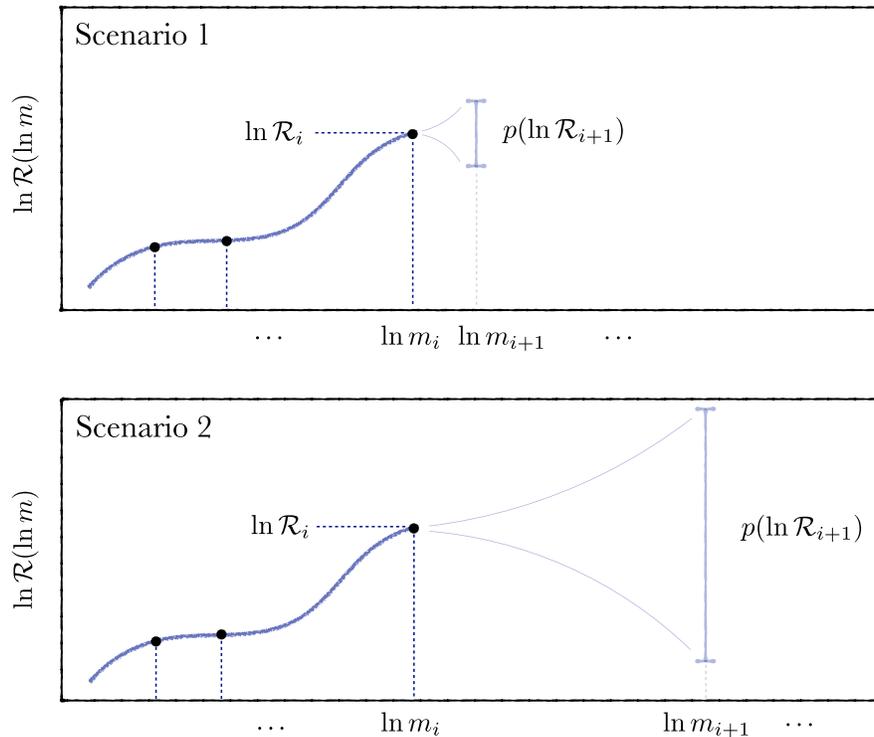

FIG. 1. Cartoon demonstrating the physical principle behind our autoregressive inference of the binary black hole population. Consider a situation in which we know the differential merger rate $\mathcal{R}(\ln m)$ at several different masses, up to some $\ln m_i$. Next, take some *new* mass, $\ln m_{i+1}$. What prior should we place on the merger rate at this new location? If $\ln m_i$ and $\ln m_{i+1}$ are close (Scenario 1), then we expect the merger rates at this location to be reasonably close as well; we might, therefore, place a tight prior on $\mathcal{R}_{i+1}$ about the value $\mathcal{R}_i$. If, on the other hand, $\ln m_i$ and $\ln m_{i+1}$ are distant (Scenario 2), the merger rates at each point are unlikely to be related, so we might place a considerably less informative prior on $\mathcal{R}_{i+1}$. This intuition is codified by choosing an autoregressive prior on $\ln \mathcal{R}(\ln m)$, such that the merger rate at any mass value has a Gaussian prior about the merger rate at the previous value, with variance related to the separation between points.

densities, considering the merger rate as a function of mass to be of the form

$$\mathcal{R}(\ln m) = r e^{\Psi(\ln m)}, \tag{5}$$

where $\Psi(\ln m)$ is an autoregressive process in $\ln m$ of order $p = 1$. This implies that, if we know the merger rate $\mathcal{R}_{i-1}$ at one mass location, then we take the rate at a new location $\ln m_i$ to be probabilistically given by the relation

$$\ln \mathcal{R}_i - \ln r = c_i[\ln \mathcal{R}_{i-1} - \ln r] + w_i, \tag{6}$$

for some choice of c_i and w_i (discussed further below). The quantity r sets the mean merger rate; it is the *departures* from $\ln r$ that are described via an AR(1) model. Note also that it is the *log* of the merger rate, not the merger rate itself, that is modeled as an AR(1) process. This choice guarantees that predicted merger rates are everywhere positive, but it has the downside that our inferred merger rate can never strictly go to $\mathcal{R} = 0$ (corresponding to $\ln \mathcal{R} \to -\infty$); see Appendix B.

We take c_i and w_i to be of the form

$$c_i = e^{-\Delta_i/\tau} \tag{7}$$

and

$$w_i = \sigma(1 - e^{-2\Delta_i/\tau})^{1/2} n_i, \tag{8}$$

where $\Delta_i = \ln m_i - \ln m_{i-1}$ is the distance between mass locations and n_i is a random variable drawn from a unit normal distribution: $n_i \sim N(0, 1)$. The parameter σ functions to rescale the random variable n_i and thus controls the allowed variance of the merger rate. The parameter τ, meanwhile, defines the mass scale over which the mass spectrum remains significantly correlated with itself. In the limit that $\Delta_i \ll \tau$, Eq. (6) demands that $\ln \mathcal{R}_i \to \ln \mathcal{R}_{i-1}$. In the opposite limit that $\Delta_i \gg \tau$, we instead have $\ln \mathcal{R}_i$ drawn randomly from $N(\ln r, \sigma)$, with no memory of earlier merger rate values. The exact forms of Eqs. (7) and (8) are chosen to ensure that σ^2 and τ indeed control the variance and autocorrelation length of the process; see Appendices A and B for more information about these expressions.

Figure 2 illustrates several random AR(1) processes $\Psi(x)$, generated with various choices of τ and σ. Processes with large τ (top panel) exhibit much stronger correlations between adjacent points, yielding larger observed scale lengths than those with smaller τ (bottom panel). Processes with large σ, meanwhile, traverse a much larger vertical range than processes with small σ. Note that these functions are continuous but do not have well-defined first derivatives. If we wanted to consider functions with continuous first derivatives, we could instead adopt "AR(2)" processes of order $p = 2$. We continue with an
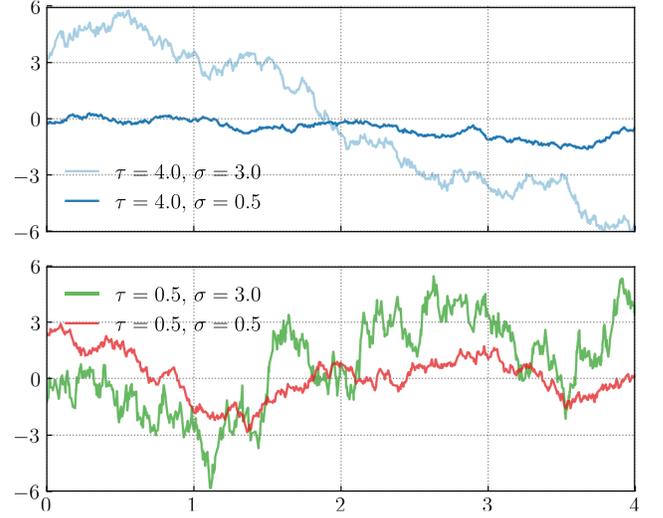

FIG. 2. Examples of various autoregressive processes $\Psi(x)$. Each curve is random draw from an AR(1) process, subject to different autocorrelation lengths τ and standard deviations σ; see Eqs. (4), (7), and (8). The top panel shows example autoregressive processes with large τ, while the bottom panel illustrates two processes with short τ. In Secs. III–VI, we model the mass, redshift, and spin distributions of binary black holes assuming they are each describable as unknown AR(1) processes.

AR(1) process, however, in order to better capture any sharp or nondifferentiable features in the binary black hole population that could be missed by models that require continuous derivatives.

B. Hierarchical inference with an autoregressive prior

Consider a set of N_{obs} gravitational-wave detections with sets of posterior samples $\{\lambda_I\}$ on the properties of each event I. The likelihood that our data, denoted $\{d\}$, arise from an underlying population described by Λ is [31–34]

$$p(\{d\}|\Lambda) \propto e^{-N_{\mathrm{exp}}(\Lambda)} \prod_{I=1}^{N_{\mathrm{obs}}} \left\langle \frac{R_d(\lambda_{I,j}; \Lambda)}{p_{\mathrm{pe}}(\lambda_{I,j})} \right\rangle_{\mathrm{samples}\ j}. \tag{9}$$

Here, the product is taken over detected events I, and the expectation value is taken over posterior samples j for each event. The quantity $p_{\mathrm{pe}}(\lambda_{I,j})$ is the prior probability assigned to each posterior sample under parameter estimation, while

$$R_d(\lambda; \Lambda) = \frac{dN}{dt_d d\lambda}(\lambda; \Lambda) \tag{10}$$

is the detector-frame merger rate density, to be evaluated at each posterior sample. We use semicolons to indicate that $R_d(\lambda; \Lambda)$ is a function of the population model Λ but not a *density* over Λ. Note also that $R_d(\lambda; \Lambda)$ is *not* a volumetric density, as in Eq. (1). If redshift z is a parameter in λ such

that R_d is a merger rate per unit redshift, then Eqs. (1) and (10) are related by

$$R_d(\lambda; \Lambda) = \mathcal{R}(\tilde{\lambda}; z, \Lambda) \frac{dV_c}{dz} (1 + z)^{-1}, \qquad (11)$$

where $\tilde{\lambda}$ is the set of all binary parameters *excluding* z and $\mathcal{R}(\tilde{\lambda}, \Lambda)$ is the volumetric merger rate density as evaluated at redshift z. The factor dV_c/dz is the differential comoving volume per unit redshift, while the factor $(1 + z)^{-1}$ is needed to convert between source-frame and detector-frame times.

Equation (9) additionally depends on $N_{\exp}(\Lambda)$, the expected number of detections over our observation time $T_{\rm obs}$ given the population Λ. We evaluate $N_{\exp}(\Lambda)$ using a set of successfully recovered signals injected into LIGO and Virgo data [6,35]. If $p_{\rm inj}(\lambda)$ is the reference probability distribution from which these injections were drawn, then [36]

$$N_{\exp}(\Lambda) \approx \frac{T_{\rm obs}}{N_{\rm inj}} \sum_{\rm found} \frac{R_d(\lambda_{{\rm inj},i}; \Lambda)}{p_{\rm inj}(\lambda_{{\rm inj},i})}, \qquad (12)$$

where $N_{\rm inj}$ is the total number of injections performed, detected or otherwise, and $T_{\rm obs}$ is our total search time. The detector-frame rate $R_d(\lambda_{{\rm inj},i}; \Lambda)$ at the location of each injection can, once again, be related to the underlying volumetric rate using Eq. (11).

The critical ingredients underlying Eqs. (9) and (12) are the differential rates $\mathcal{R}(\lambda_{I,j}; z, \Lambda)$ and $\mathcal{R}(\lambda_{{\rm inj},i}; z, \Lambda)$ at the locations of every posterior sample and every found injection. In the usual strongly parametrized approach, we obtain these quantities by assuming some functional form for $\mathcal{R}(\lambda; z, \Lambda)$. Here, our goal is to not assume a particular functional form for the differential rate, but to directly infer the merger rate at every posterior sample and every found injection using our autoregressive prior. In adopting this approach, we have rid ourselves of (nearly) all ordinary hyperparameters. Instead, the merger rates at every posterior sample and every injection are themselves the parameters that we directly infer from the data. This results in a rather large-dimensional parameter space. If we have $N_{\rm obs}$ events (each with $N_{\rm samp}$ posterior samples) and $N_{\rm inj}$ injections, we are directly inferring the binary merger rate at $N_{\rm obs}N_{\rm samp} + N_{\rm inj}$ discrete locations. The form of Eq. (6), however, imposes an almost equally large number of constraints, ensuring that the inference problem remains tractable.

We have not quite discarded all hyperparameters: We still need to determine the variance σ and autocorrelation length τ associated with our autoregressive rate prior. Rather than fix σ and τ, we hierarchically infer them as well, allowing our data to dictate the characteristic length scale and size of features present in the binary black hole

population. In Appendix B, we derive and discuss the priors we place on σ and τ. To obtain physically meaningful priors, we approach the problem indirectly, considering not constraints on σ and τ themselves but instead on allowed variations in the black hole merger rate $\mathcal{R}(\lambda)$; these choices then induce priors on σ, τ, and the ratio $\sigma/\sqrt{\tau}$.

It is worthwhile to compare our methodology to other flexible approaches appearing in the literature. One similar approach is the spline-based method appearing in Refs. [19–21]; this model proceeds by first defining merger rates over a discrete set of "knot" locations and then constructing a spline interpolant between these knots. The rates at each knot location may themselves be linked via a Gaussian process prior or other regularization schemes [20]. The "binned Gaussian process" models in Refs. [6,22–26] operate similarly. In this approach, a merger rate is again defined across a discrete grid of points and interpolated, but now assuming that the merger rate is a piecewise constant between grid points rather than a spline.

A primary methodological difference between these approaches and ours is that we do not perform interpolation: The parameters governing our models are the direct merger rates at each point of interest rather than the rates defined over some reference grid. Our autoregressive model also behaves in ways that make it complementary to these other approaches. Because spline interpolants are continuous in their derivatives, the spline-based approaches above are suitable for identifying smooth trends in the data but may struggle to resolve sharp features or features at the same scale as the knot separations. The continuity and differentiability imposed by spline models can additionally sometimes give rise to oscillatory "ringing" that depends on the precise choice of knot locations [21,37]. Our AR(1) model requires no differentiability, however, avoiding this oscillatory behavior. The lack of a reference grid also means that we require no *a priori* choice of scale. This freedom, however, greatly boosts the computational cost of our approach and gives rise to possible instabilities in the hierarchical likelihood; this instability is described in Appendix B. Even with the flexibility afforded by an AR(1) process, there remain limitations on the degree to which our model can recover discontinuously sharp features; see Appendix C for further discussion.

We implement our autoregressive model using jax [38] and numpyro [39,40], which enable compilation and autodifferentiation of our likelihood. We perform our Bayesian inference using numpyro's implementation of the NUTS ("No U-Turn Sampler") algorithm [41], a variant of Hamiltonian Monte Carlo (HMC) sampling [42]. As noted above, our autoregressive models actually comprise a vast number of latent parameters: one per posterior sample and found injection. In practice, this amounts to approximately 2.5×10^5 parameters for the analyses presented in this paper. Given this extremely high-dimensional space, the computational acceleration and sampling efficiency

afforded by autodifferentiation and HMC methods is critical. Further details regarding our hierarchical inference, including the exact data and priors used, are given in Appendix B.

III. STOP ONE: MASSES

We first use our autoregressive model to investigate the distribution of binary black hole primary masses m_1 and mass ratios q. We consider the merger rate to be the combination of two parallel autoregressive processes, $\Psi(\ln m_1)$ and $\Phi(q)$, that capture the dependence of the merger rate on both $\ln m_1$ and q:

$$\mathcal{R}(\ln m_1, q, \chi_1, \chi_2, \cos\theta_1, \cos\theta_2; z)$$
$$= r[e^{\Psi(\ln m_1)}e^{\Phi(q)}]\left(\frac{1+z}{1+0.2}\right)^{\kappa}p(\chi_1, \chi_2, \cos\theta_1, \cos\theta_2).$$
(13)

We fit for both $\Psi(\ln m_1)$ and $\Phi(q)$ simultaneously, allowing each process to possess its own variance and autocorrelation length.

While our focus in this section is on the mass distribution of binary black holes, when measuring the population distribution of any one parameter it is usually important to simultaneously fit the distributions of other parameters, like spin magnitudes χ_i, spin-orbit misalignment angles θ_i, and redshifts. There is no fundamental reason why the distributions over *all* binary parameters cannot be fit as simultaneous autoregressive processes. Since each additional AR(1) process introduces a fairly high computational cost, however, for simplicity we will fit the "leftover" redshift and spin distributions by falling back on ordinary parametrized models. We assume that the merger rate evolves as $(1+z)^{\kappa}$ for some unknown index κ [43] and that component spins are independently and identically distributed with a probability distribution $p(\chi_1, \chi_2, \cos\theta_1, \cos\theta_2)$ of the form given in Appendix D; these redshift and spin distributions are hierarchically fit alongside our autoregressive mass model. Finally, note that our model in Eq. (13) is presumed to be separable, with no correlations between the masses, mass ratios, and spins of binary black holes.

A. Features in the black hole mass distribution

The top panel of Fig. 3 shows our autoregressive measurement on the merger density rate of binary black holes as a function of primary mass, evaluated at $q = 1$ and $z = 0.2$ and marginalized over spin degrees of freedom. Each blue trace shows a single posterior sample for $\mathcal{R}(\ln m_1)$, [44] while the thick and thin black curves mark a running median and central 90% credible bounds, respectively. We note that our presentation of the mass spectrum, conditioned on some particular reference values

of mass ratio and redshift, is slightly unusual; it is more common to show a mass distribution that has been fully marginalized over other parameters. When marginalizing a merger rate over one or more parameters, however, the result can show extreme systematic dependence on the exact model presumed for these marginalized parameters, particularly across regions of parameter space that are not well measured. An extreme example can be found in Ref. [6], in which the fully marginalized binary neutron star merger rate can vary by 2 orders of magnitude depending on the mass model used. Our approach in this paper is to minimize such systematics by instead quoting differential merger rates at well-measured locations in parameter space (e.g., $q = 1$ and $z = 0.2$); this approach maximizes precision and best enables comparison to predictions between observation and theory.

Returning to Fig. 3, we see three possible features in the black hole primary mass spectrum:

(1) *A global maximum at* $m_1 \approx 10M_{\odot}$. The binary merger rate appears to be maximized at $m_1 \approx 10M_{\odot}$ primary masses, falling off with both lower and higher primary masses. We can quantify the significance of this feature by computing the fraction of posterior samples that exhibit a systematic peak in this neighborhood. To do so, we compute and compare the average merger rates across three bins: $7.5–9M_{\odot}$, $9–11M_{\odot}$, and $11–13.5M_{\odot}$ (chosen to have roughly equal logarithmic widths). We regard a "peak" as a case when the averaged merger rate in the middle interval is higher than the averaged merger rates in both adjacent bins. As shown in the top panel of Fig. 4, we find that 96% of our samples meet this criterion and exhibit a systematic peak near $10M_{\odot}$.

(2) *A local maximum at* $m_1 \approx 35M_{\odot}$. We can again quantify the significance of this feature by comparing the average rates across three bins: $20–28M_{\odot}$, $28–40M_{\odot}$, and $40–55M_{\odot}$. As shown in the middle panel of Fig. 4, 94% of our posterior draws yield higher averaged merger rates in the $28–40M_{\odot}$ range than in both adjacent bins. Thus, both the $10M_{\odot}$ and $35M_{\odot}$ maxima have roughly equal significance; although neither is unambiguously required by the data, both are favored to exist at greater than 90% credibility.

(3) *Steepening of the continuum above* $40M_{\odot}$. Between the $10M_{\odot}$ and $35M_{\odot}$ maxima is a large, relatively flat continuum. Above the $35M_{\odot}$ maximum, the continuum appears to steepen, falling off more rapidly with increasing mass. We quantify the evidence for this steepening by computing and comparing the mean power-law slope of the black hole merger rate above and below the $35M_{\odot}$ maximum. From each posterior sample, we extract the merger rates $\mathcal{R}(\ln m_1)$ near 15, 25, 45, and $85M_{\odot}$; these rates are then used to compute
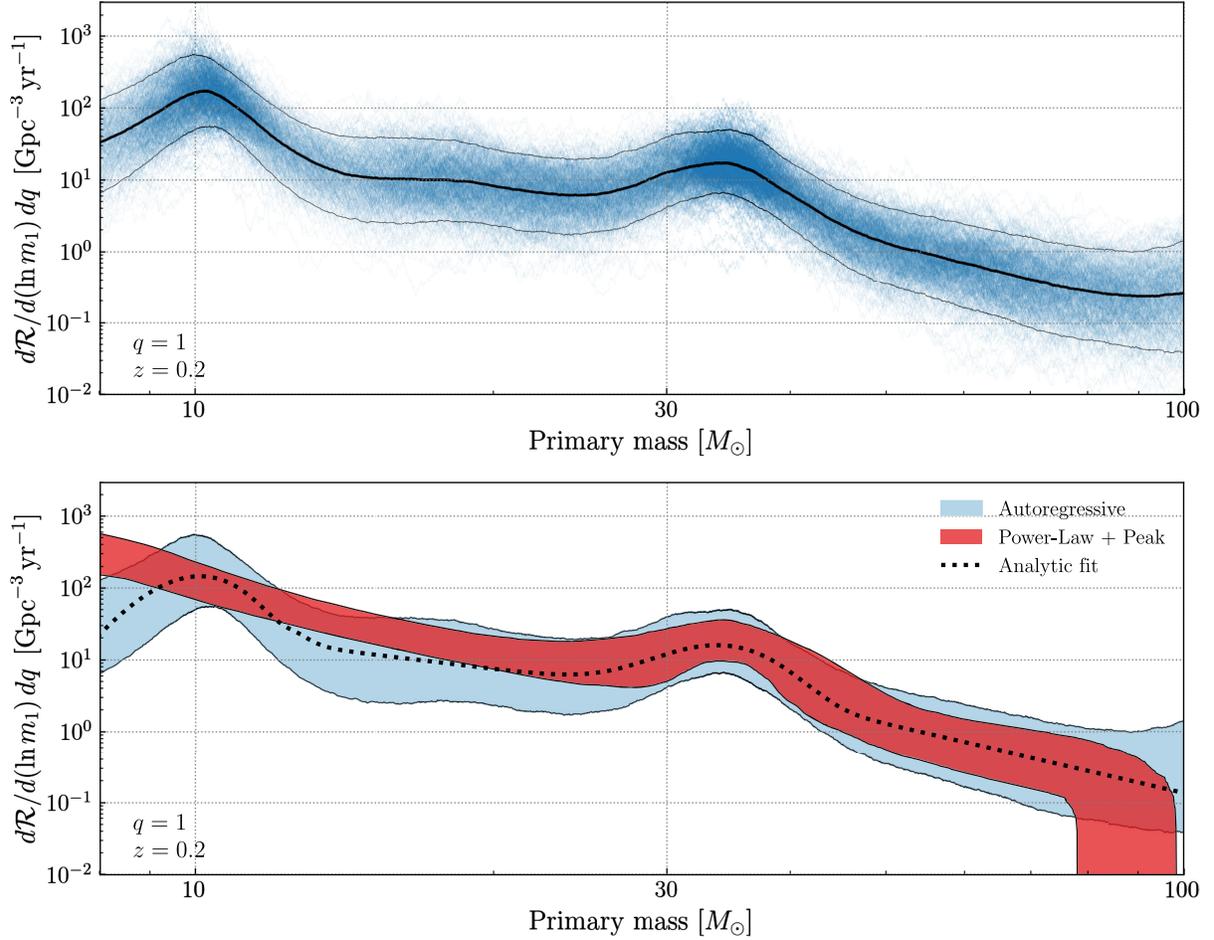
FIG. 3. Top panel: binary black hole merger rate as a function of primary mass, inferred nonparametrically under an autoregressive prior. The merger rate is evaluated at mass ratio $q = 1$ and redshift $z = 0.2$, and marginalized over spins, following the model defined in Eq. (13). The solid black trace marks the mean inferred $\mathcal{R}(\ln m_1)$ as a function of m_1, while the lighter black traces bound our central 90% credible bounds. Individual posterior draws on $\mathcal{R}(\ln m_1)$ are shown via light blue traces. Three features naturally emerge in the inferred mass distribution: a global maximum in the merger rate at $m_1 \approx 10 M_\odot$, a secondary maximum at $m_1 \approx 35 M_\odot$, and an otherwise smooth continuum that steepens above $40 M_\odot$. Each of these three features is exhibited by approximately 90% of posterior draws. Bottom panel: comparison between our autoregressive inference (blue band) and results obtained using the strongly parametrized POWERLAW+PEAK model in Ref. [6] (red). Rates are again evaluated at $q = 1$ and $z = 0.2$, and marginalized over spin. Each band encompasses the central 90% credible region inferred using the given model. Both approaches give consistent estimates of the merger rate at $m_1 \approx 10 M_\odot$, as well as the merger rate in the $30 - 70 M_\odot$ interval. In order to match these rates, though, we see that the parametrized model is forced to overestimate the merger rate between 15 and $30 M_\odot$, as well as the merger rate below $10 M_\odot$. Furthermore, our autoregressive model shows no indication of a sharp cutoff in the binary mass distribution at or above $80 M_\odot$ (this feature is included *a priori* in the strongly parametrized model). A simple fit to our median inferred rate, using the parametric form of Eq. (16), is shown via the black dotted curve.

the power-law indices characterizing the middle and high end of the mass spectrum:

$$\alpha_{\mathrm{mid}} = \frac{\ln \overline{\mathcal{R}(25 M_\odot)} - \ln \overline{\mathcal{R}(15 M_\odot)}}{\ln(25 M_\odot) - \ln(15 M_\odot)} \quad (14)$$

and

$$\alpha_{\mathrm{high}} = \frac{\ln \overline{\mathcal{R}(85 M_\odot)} - \ln \overline{\mathcal{R}(45 M_\odot)}}{\ln(85 M_\odot) - \ln(45 M_\odot)}. \quad (15)$$

We write $\overline{\mathcal{R}(25 M_\odot)}$, for example, to indicate the average merger rate in a $1 M_\odot$ window around $25 M_\odot$. Using window-averaged rates in this fashion enables more reliable estimates of representative power-law indices due to the rapid oscillations exhibited by individual $\mathcal{R}(\ln m_1)$ traces. The joint distribution of both power-law slopes is plotted in the lower panel of Fig. 4. In the 15–$25 M_\odot$ interval, we find an average power-law index $\alpha_{\mathrm{mid}} = -1.1^{+2.7}_{-2.6}$, while in the 45–$85 M_\odot$ range, we find $\alpha_{\mathrm{high}} = -3.8^{+2.6}_{-2.7}$. We identify a preference for steepening,
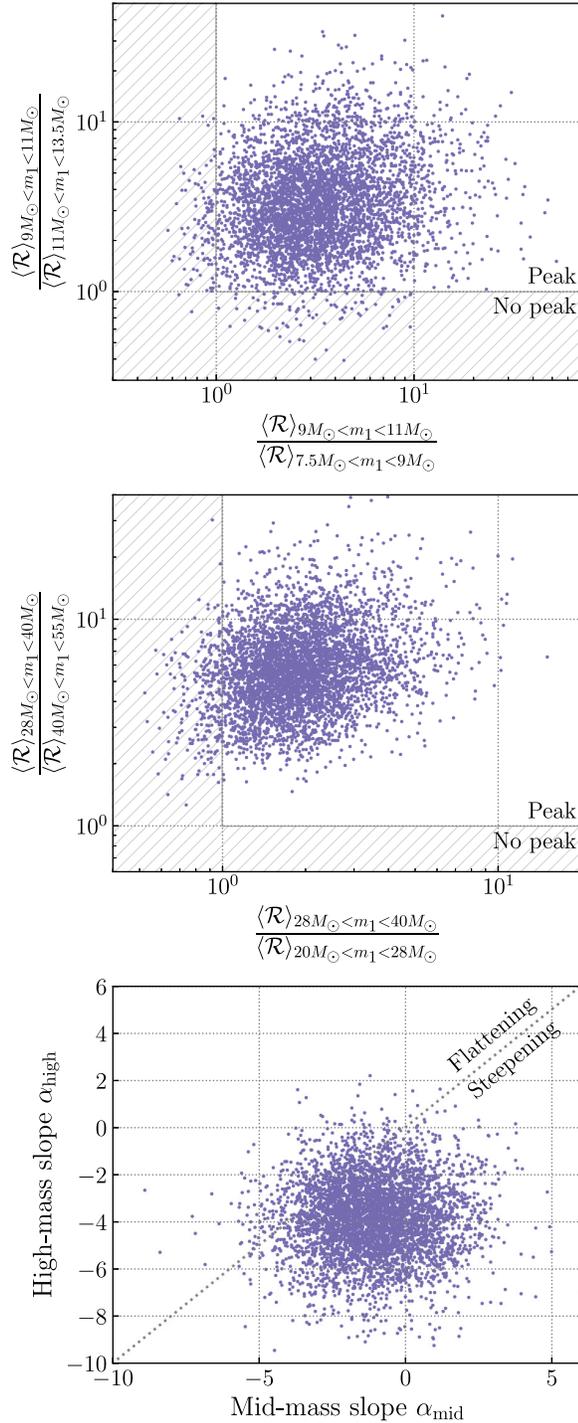
FIG. 4. Tests quantifying the significance of features identified in Fig. 3. Top panel: ratios between the average merger rate across $9M_\odot < m_1 < 11M_\odot$ and in adjacent lower- and higher-mass intervals. If a peak is present near $10M_\odot$, both ratios should be greater than 1; this is true for 96% of our samples. Middle panel: similarly, ratios between the average merger rate across $28M_\odot < m_1 < 40M_\odot$ and in adjacent bands. A peak near $35M_\odot$ is present in 94% of samples. Bottom: implied power-law indices characterizing the 15–25$M_\odot$ and 45–75$M_\odot$ intervals (α_{mid} and α_{high}, respectively). We find that 89% of samples show a steepening in the mass spectrum, with $\alpha_{high} < \alpha_{mid}$.

with $\alpha_{high} < \alpha_{mid}$, in 89% of samples, although this behavior is not strictly required by the data.

B. Discussion

The significances of the $10M_\odot$ and $35M_\odot$ peaks, as computed here, are similar to but more conservative than significance estimates presented elsewhere. A strongly parameterized analysis presented in Ref. [6] identifies a $35M_\odot$ excess at effectively 100% credibility, and an analysis in the same study using splines to measure deviations from an ordinary power law finds upward fluctuations at $10M_\odot$ and $35M_\odot$ with greater than 99% credibility (see also Refs. [19,20]). Reference [37] alternatively explores the frequency with which apparent peaks might arise purely from random counting statistics due to our still-moderate number of binary black hole detections. By repeatedly drawing realizations of 69 events from a peakless power-law population, that study finds the observed $10M_\odot$ and $35M_\odot$ peaks to be more statistically significant than over 99% of false peaks arising from random clustering.

The difference between these significance estimates and ours is likely twofold. First, these significance estimates test slightly different features; an upward fluctuation relative to a power law does not necessarily indicate a local maximum but can also be caused by a plateau or change in slope. Second, by virtue of its extreme flexibility, our autoregressive prior likely maximizes the variance in our $\mathcal{R}(\ln m_1)$ measurements, slightly diminishing our confidence in any given feature. We note that our assessment of feature significance does not depend on the particular choice of reference mass ratio and redshift adopted in Fig. 3; different reference values would rescale each merger rate by a hyperparameter-dependent constant, which cancels when subsequently taking ratios between rates as in Fig. 4.

In addition to the $10M_\odot$ and $35M_\odot$ maxima, other studies have noted the possible existence of other features in the primary mass spectrum, namely, additional maxima or minima in the 15–25$M_\odot$ range [6,13,20,45]. We do not see evidence for any such features here, however, indicating that any additional features are likely prior dependent and consistent with random clustering of a still-small number of observations. References [6,19,20,37] note a somewhat significant *dip* in the mass spectrum, relative to a power law, near $14M_\odot$. We interpret this result not as a local minimum but just as a flattening of the power-law index at lower masses, as seen in Fig. 3 and discussed further below. Additionally, various studies have searched for the presence of a high-mass cutoff in the black hole mass spectrum [6,9,10,12,24,25,46–48], possibly due to the occurrence of pair-instability supernova [49]. In Ref. [6], for example, it is inferred that if such a cutoff exists, then it must occur at $m_1 > 78M_\odot$ at 95% credibility. Our analysis, however, shows no indication of a cutoff in the black hole mass

spectrum, instead recovering a distribution that continues to smoothly decline out to $m_1 \approx 100 M_\odot$. Note that the slight increase in the variance of $\mathcal{R}$ seen near $100 M_\odot$ marks a reversion to the prior in the region $m_1 \gtrsim 100 M_\odot$ where we have little data.

It is valuable to compare our autoregressive results to measurements made using standard strongly parametrized models, in order to identify regions where strongly parametrized models may fail to capture features in the data and to guide the iterative development of improved models going forward. In the bottom panel of Fig. 3, we compare our autoregressive model of the binary merger rate (blue) with results obtained under the POWERLAW+PEAK model [10] presented in Ref. [6]. Both models identify an excess of mergers near $35 M_\odot$, and both measure approximately consistent merger rates near $10 M_\odot$. We see two signs of tension, however. First, the POWERLAW+PEAK model otherwise adopts a single unbroken power law; in order to match the merger rate at both low $(10 M_\odot)$ and high $(\geq 30 M_\odot)$ masses, it is therefore forced to overestimate the merger rate in the 15–$30 M_\odot$ range. This result is consistent with the *downward* perturbation identified by spline-based methods in this region [6,19,20,37]; this downward perturbation may not be caused by a local minima in the mass spectrum but just a flattening of the power-law index at lower masses.

C. Suitable parametric model

In cases where a strongly parametrized phenomenological model is needed, our autoregressive result suggests that a sufficient choice is a model comprising two Gaussian peaks and a broken power law, with a probability density

$$p(m_1) = f_{p,1} N(m_1 | \mu_{m,1}, \sigma_{m,1}) + f_{p,2} N(m_2 | \mu_{m,2}, \sigma_{m,2})$$
$$+ (1 - f_{p,1} - f_{p,2}) \Gamma(m_1). \quad (16)$$

Here, we use $N(m_1 | \mu, \sigma)$ to signify a normalized Gaussian distribution with mean μ and standard deviation σ, and $\Gamma(m_1)$ to denote a broken power law tapered towards zero at low masses:

$$\Gamma(m_1) \propto \begin{cases} e^{-\frac{(m_1 - m_{\min})^2}{2\delta m^2}} \left(\frac{m_1}{m_b}\right)^{\alpha_1} & (m_1 < m_{\min}) \\ \left(\frac{m_1}{m_b}\right)^{\alpha_1} & (m_{\min} \leq m_1 < m_b) \\ \left(\frac{m_1}{m_b}\right)^{\alpha_2} & (m_b \leq m_1 < m_{\max}) \\ 0 & (\text{else}), \end{cases} \quad (17)$$

with a proportionality constant chosen to enforce $\int \Gamma(m_1) dm_1 = 1$. A least-squares fit against our mean inferred $\ln \mathcal{R}(\ln m)$ gives best-fit parameters

$$\mu_1 = 10.0 M_\odot, \qquad \mu_2 = 33.4 M_\odot,$$
$$\sigma_1 = 1.1 M_\odot, \qquad \sigma_2 = 4.2 M_\odot,$$
$$f_{p,1} = 0.75, \qquad f_{p,2} = 0.07,$$
$$\alpha_1 = -2.7, \qquad \alpha_2 = -4.2,$$
$$m_{\min} = 12.9 M_\odot, \qquad \delta_m = 0.6 M_\odot,$$
$$m_b = 42.7 M_\odot, \qquad m_{\max} = 100 M_\odot. \quad (18)$$

The corresponding distribution $p(\ln m_1) = p(m_1) m_1$ is shown as a dotted line in Fig. 3. Note that this fit approximates the fully marginalized primary mass distribution and is thus valid at any choice of q, z, etc.

D. Features in the black hole mass-ratio distribution

Compared to the primary mass distribution, we resolve relatively little information about the distribution of black hole mass ratios. The top panel of Fig. 5 illustrates our constraints on $\mathcal{R}(q)$, evaluated at $z = 0.2$ and $m_1 = 20 M_\odot$, and integrated over component spins. The only feature that manifests in Fig. 5 is a possible preference for larger q. As above, we can compare integrated merger rates in two bands, $0.5 \leq q \leq 0.6$ and $0.9 \leq q \leq 1$, to quantify the significance of this feature. We find that the merger rate in the high-q interval is greater than the rate in the low-q interval for 90% of samples, such that the binary black hole population likely favors equal mass ratios.

In the lower panel of Fig. 5, we compare our results with the strongly parametrized measurements presented in Ref. [6] using the POWERLAW+PEAK model, in which the mass ratio distribution is modeled as a power law with a primary-mass-dependent truncation:

$$p(q|m_1) \propto S(q; m_1) q^{\beta_q}. \quad (19)$$

Here, $S(q; m_1)$ is a tapering function that sends $p(q|m_1)$ to zero when $q < m_{\min}/m_1$ for some $m_{\min}$. Both results are again evaluated at $z = 0.2$ and $m_1 = 20 M_\odot$, and integrated over black hole spins. Other than the truncation below $q \approx 0.2$ (imposed in Ref. [6] as an *a priori* modeling choice), both sets of results are broadly consistent. In the strongly parametrized analysis of Ref. [6], it is found that $\beta_q > 0$ with 92% credibility, comparable to our significance estimate above [50].

IV. STOP TWO: REDSHIFTS

Next, we investigate the redshift distribution of binary black holes. In most analyses, the redshift dependence of the binary black hole merger rate is presumed to follow a power-law form: $\mathcal{R}(z) \propto (1 + z)^\kappa$ for some index κ [6,9,43,48,51,52]. Under this model, it has been concluded that the binary black hole merger rate systematically grows with redshift at a rate consistent with star formation in the local Universe [6]. Here, we instead model the redshift
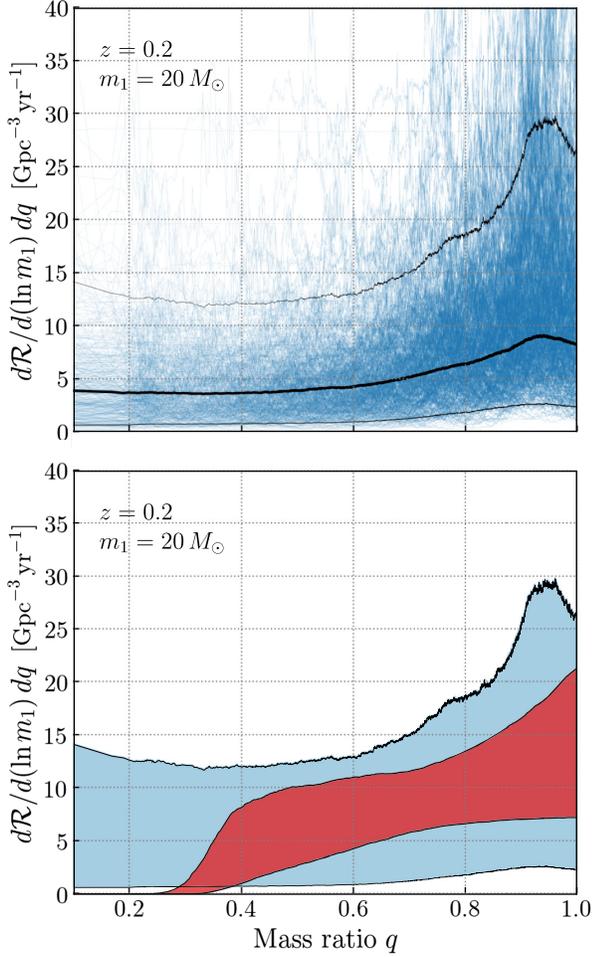
FIG. 5.   Top panel: merger rate of binary black holes as a function of mass ratio, evaluated at $m_1 = 20 M_\odot$ and $z = 0.2$, and integrated over possible spins, following Eq. (13). The thick and thin black lines mark the mean and central 90% bounds on $\mathcal{R}(q)$, while thin blue traces show individual draws from our posterior on $\mathcal{R}(q)$. We see a preference for an increasing merger rate as a function of q, but this behavior is not strictly required. Bottom panel: comparison between $\mathcal{R}(q)$ as inferred by our autoregressive model (blue) and the strongly parametrized analysis of Ref. [6] (red), which assumes a power-law dependence on q with a truncation in the merger rate below $q = m_{\min}/m_1$ for some minimum mass $m_{\min}$. The two results are broadly consistent, although under our autoregressive model, we find reduced evidence for a merger rate that increases with larger q.

dependence of the black hole merger rate as an autoregressive process, searching for any features that might be missed under a more strongly parametrized approach. We simultaneously measure the mass and component spin distributions by falling back on the strongly parametrized models described in Appendix D. Together, our model is of the form

$$\mathcal{R}(\ln m_1, q, \chi_1, \chi_2, \cos\theta_1, \cos\theta_2; z)$$
$$= r\frac{f(m_1, q)}{f(20 M_\odot, 1)} e^{\Psi}(z) p(\chi_1, \chi_2, \cos\theta_1, \cos\theta_2). \quad (20)$$

A. Features in the black hole redshift distribution

The left panel of Fig. 6 shows our resulting inference on the binary merger rate as a function of redshift, evaluated at $m_1 = 20 M_\odot$ and $q = 1$ and integrated across spins. Blue traces show individual draws from our posterior, the solid black curve marks the running median rate, and thin gray lines denote central 90% credible bounds on the merger rate at each redshift. The right panel of Fig. 6 compares these results (in blue) to the results obtained in Ref. [6] using the strongly parametrized power-law model for the black hole merger rate. Both approaches yield consistent estimates of the merger rate at $z \approx 0.3$ and $z \approx 1$, but our autoregressive result suggests that the intervening evolution is not necessarily well modeled by a power law. Instead, our result is consistent with a "sigmoid" shape displaying the following features:

(1) *A nonevolving, uniform-in-comoving-volume rate below $z \approx 0.4$.* At the lowest redshifts, the data do not require the merger rate to evolve with redshift. Instead, our autoregressive results are consistent with a rate that remains constant out to $z \approx 0.4$. To gauge the significance of this feature, we compute and compare the mean merger rates in two intervals: $0.1 < z < 0.2$ and $0.3 < z < 0.4$. As shown in the left panel of Fig. 7, we find these mean rates to be consistent with one another, with the mean rate in the $0.3 < z < 0.4$ interval exceeding the rate in the $0.1 < z < 0.2$ interval only 53% of the time.

(2) *A rise in the merger rate between $z \approx 0.4$ and 0.8.* Beyond redshift $z \approx 0.4$, however, the merger rate is required to increase by up to an order of magnitude by $z \approx 0.8$. We quantify the significance of this rise by comparing the mean merger rate in the range $0.1 < z < 0.2$ to the mean rate in the range $0.7 < z < 0.8$. As shown in the right panel of Fig. 7, the mean rates in these high- and low-redshift intervals are confidently unequal, with the $0.7 < z < 0.8$ merger rate exceeding the $0.1 < z < 0.2$ rate 93% of the time. Beyond redshift $z \approx 1$, the absence of informative data causes our measurement to asymptote back towards the autoregressive prior, yielding expanding error bars towards higher redshifts.

B. Discussion

Other studies employing flexible nonparameteric analyses have also obtained results indicating a possible tension with a $(1 + z)^\kappa$ power law. Reference [53] explored the use of population models composed of "Green's-function"-like delta functions as a tool with which to diagnose the performance of strongly parametrized models. In that work, the likelihood is found to be maximized when $\mathcal{R}(z)$ is modeled as a sequence of delta functions that initially *decrease* in height below $z \approx 0.13$, followed by an elevated
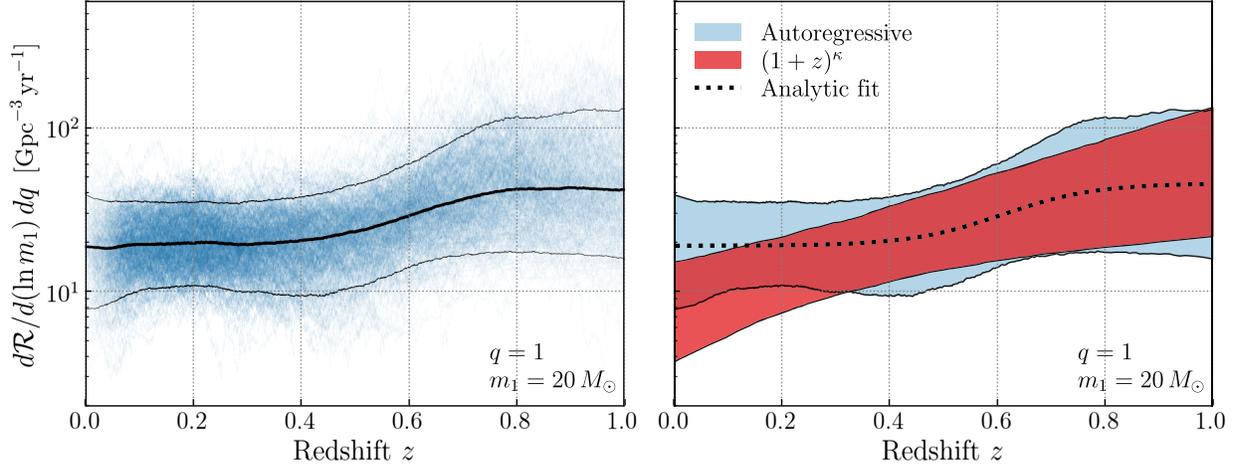
FIG. 6. Left panel: binary black hole merger rate as a function of redshift, inferred nonparametrically using an autoregressive process prior. The merger rate is evaluated at a primary mass $m_1 = 20 M_\odot$ and mass ratio $q = 1$, and integrated over black hole spins. Light blue traces show individual draws from our posterior, while the black and gray curves denote a running median and central 90% credible bounds, respectively. Right panel: comparison between our nonparametric result (blue) and the result obtained in Ref. [6] when assuming that the merger rate evolves as $(1 + z)^\kappa$ for an unknown index κ. Both bands denote 90% credible bounds. We see that both approaches recover similar merger rates at $z \approx 0.3$ and $z \approx 1$, and both indicate that the black hole merger rate systematically grows with redshift. Our autoregressive result, however, suggests that this growth may not be well modeled by a power law but instead by a slowly growing or constant merger rate that begins to evolve more sharply only beyond $z \gtrsim 0.4$. The dashed black curve, for example, shows the result of a simple least-squares fit to our median inferred merger rate using the sigmoid model defined in Eq. (22). A broken power law, as in Eq. (21), also yields a good fit at $z \lesssim 0.8$.
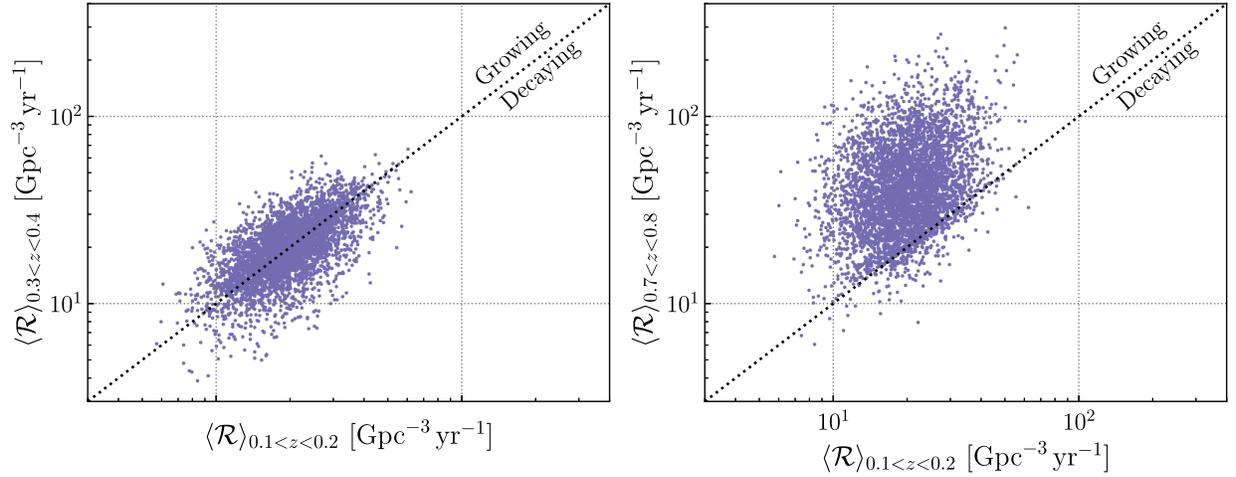

FIG. 7. Left panel: comparison between the mean merger rate across the interval $0.3 < z < 0.4$ and the mean rate across $0.1 < z < 0.2$. Each point corresponds to a single posterior draw from Fig. 6. All estimates cluster around the diagonal, indicating that the merger rates in both intervals are consistent with one another. The data are therefore consistent with a nonevolving merger rate below $z \lesssim 0.4$. Right panel: analogous comparison between the mean merger rate in the interval $0.7 < z < 0.8$ and the mean rate within $0.1 < z < 0.2$. The merger rate in the high-redshift interval is greater than that in the low-redshift interval for 96% of samples, indicating a preference for a merger rate that grows at large redshifts.

but flat merger rate in the range $0.2 \lesssim z \lesssim 0.5$ that then more sharply rises between $0.5 \lesssim z \lesssim 0.75$; see their Fig. 5. Other than the initially decreasing merger rate, which we do not recover, these results are consistent with the behavior we see in Fig. 6. Reference [20], in turn, measured the redshift-dependent merger rate using a set of basis splines to capture deviations from a $(1 + z)^\kappa$ power law. That work

also recovered a largely constant merger rate density below $z \approx 0.4$, followed by a steeper increase in the merger rate out to $z \approx 1$; see their Fig. 8.

If real, the steplike structure in the redshift-dependent merger rate could arise from a variety of effects. The redshift-dependent merger rate $\mathcal{R}(z)$ is generally modeled by convolving an estimate of the metallicity-dependent

cosmic star formation rate with a distribution of time delays between progenitor formation and binary merger; the time delay distribution is itself typically modeled as a power law. The resulting merger rate is also usually well described by a power law at low redshifts. If the observed binary black hole population is dominated by a single formation channel, the possible non-power-law behavior in Fig. 6 could indicate additional nontrivial structure in the birth rate or time delay distribution of binary progenitors. Alternatively, the observed binary population could comprise a mixture of several distinct formation channels. A shift from a flat to an evolving merger rate at $z \approx 0.5$ could mark a transition between two formation channels, one of which dominates low-redshift mergers and the other of which takes over at larger redshifts. If a mixture between formation channels is the correct explanation of Fig. 6, then we should also expect to see systematic evolution in *other* intrinsic properties of binary black holes between low and high redshifts. Although no such evolution has been found in the binary black hole mass spectrum [48,52], the binary black hole *spin* distribution potentially evolves with redshift, with the effective inspiral spin (further discussed in Sec. VI below) becoming larger and more positive at higher z [54]. Additional observations will be critical in confirming the trends identified in Fig. 6 and in Ref. [54] and in probing any relationship between these two trends.

C. Suitable parametric models

When a parametric model is required, our autoregressive results suggest that one might replace the standard power-law model with a broken power law:

$$\frac{d\mathcal{R}}{d \ln m_1 dq}(z) = \begin{cases} \mathcal{R}_b \left(\frac{1+z}{1+z_b}\right)^{\kappa_1} & (z \le z_b) \\ \mathcal{R}_b \left(\frac{1+z}{1+z_b}\right)^{\kappa_2} & (z > z_b), \end{cases} \quad (21)$$

with a transition between power-law indices κ_1 and κ_2 occurring at $z = z_b$, or a sigmoid,

$$\frac{d\mathcal{R}}{d \ln m_1 dq}(z) = \mathcal{R}_0 + \frac{\delta\mathcal{R}}{1 + e^{-(z-z_b)/\delta z}}, \quad (22)$$

in which the merger rate density increases from $\mathcal{R}_0$ to $\mathcal{R}_0 + \delta\mathcal{R}$ across an interval of width δz around a transition redshift z_b. A least-squares fit to our median $\ln \mathcal{R}$ using Eq. (21) gives

$$\begin{aligned} \kappa_1 &= 0.3, \\ \kappa_2 &= 3.4, \\ z_b &= 0.44, \\ \mathcal{R}_b &= 20.4 \text{ Gpc}^{-3} \text{ yr}^{-1}. \end{aligned} \quad (23)$$

A fit using Eq. (22), in turn, gives

$$\begin{aligned} z_b &= 0.65, \\ \delta z &= 0.08, \\ \mathcal{R}_0 &= 19.1 \text{ Gpc}^{-3} \text{ yr}^{-1}, \\ \delta\mathcal{R} &= 27.0 \text{ Gpc}^{-3} \text{ yr}^{-1}; \end{aligned} \quad (24)$$

this fit is shown as a dotted curve in Fig. 6. As our autoregressive results begin to revert to the prior above $z = 1$, these fits are performed only in the restricted range $z \le 0.8$.

Recall that the above fits describe the merger rate per $\ln m_1$ per unit q evaluated at $m_1 = 20M_\odot$ and $q = 1$, rather than the fully integrated merger rate. If the full binary black hole merger rate, integrated over all masses, is desired, it can be fit with the same functional forms:

$$\mathcal{R}(z) = \begin{cases} \mathcal{R}_b \left(\frac{1+z}{1+z_b}\right)^{\kappa_1} & (z \le z_b) \\ \mathcal{R}_b \left(\frac{1+z}{1+z_b}\right)^{\kappa_2} & (z > z_b), \end{cases} \quad (25)$$

or

$$\mathcal{R}(z) = \mathcal{R}_0 + \frac{\delta\mathcal{R}}{1 + e^{-(z-z_b)/\delta z}}, \quad (26)$$

with parameters

$$\begin{aligned} \kappa_1 &= 0.2, \\ \kappa_2 &= 3.5, \\ z_b &= 0.44, \\ \mathcal{R}_b &= 26.4 \text{ Gpc}^{-3} \text{ yr}^{-1} \end{aligned} \quad (27)$$

in Eq. (25) or

$$\begin{aligned} z_b &= 0.64, \\ \delta z &= 0.08, \\ \mathcal{R}_0 &= 25.0 \text{ Gpc}^{-3} \text{ yr}^{-1}, \\ \delta\mathcal{R} &= 33.9 \text{ Gpc}^{-3} \text{ yr}^{-1} \end{aligned} \quad (28)$$

in Eq. (26).

V. STOP THREE: COMPONENT SPINS

Next, we turn to the distribution of spins among binary black hole systems. A black hole binary is characterized by 6 spin degrees of freedom, 3 per component spin. Assuming that component spins have no preferential azimuthal orientations (although see Ref. [55]), we work in a reduced four-dimensional space and fit for the distributions of component spin magnitudes χ_1 and χ_2 and (cosine of the) spin-orbit tilt angles $\cos\theta_1$ and $\cos\theta_2$. We assume that the

variation of the merger rate across spin magnitudes and tilts is described via two autoregressive processes, $\Psi(\chi)$ and $\Phi(\cos\theta)$, with the two component spins in a given binary distributed independently and identically. As we measure $\Psi(\chi)$ and $\Phi(\cos\theta)$, we simultaneously infer the mass and redshift distributions of the binary black hole population by falling back on ordinary strongly parametrized models, assuming a primary mass and mass ratio distributions $f(m_1)$ and $p(q)$ as described in Appendix D and a merger rate density that grows as $(1+z)^\kappa$. Together, our full merger rate model is of the form

$$\mathcal{R}(\ln m_1, q, \chi_1, \chi_2, \cos\theta_1, \cos\theta_2; z)$$
$$= r\frac{f(m_1)p(q)}{f(20M_\odot)}\left(\frac{1+z}{1+0.2}\right)^\kappa [e^{\Psi(\chi_1)}e^{\Psi(\chi_2)}e^{\Phi(\cos\theta_1)}e^{\Phi(\cos\theta_2)}].$$
(29)

A. Features in the black hole spin distribution

Figure 8 shows our autoregressive measurements of the black hole spin magnitude and tilt distributions. We plot our results in two ways. First, the upper row shows merger rates as a function of spin magnitude and orientation. The upper left panel shows the merger rate of binaries along the $\chi_1 = \chi_2 = \chi$ diagonal at fixed reference mass, mass ratio, redshift, and spin tilts ($m_1 = 20M_\odot$, $q = 1$, $z = 0.2$, and $\cos\theta_1 = \cos\theta_2 = 1$); using Eq. (29), this is given by

$$\mathcal{R}(\chi_1, \chi_2 = \chi) = r(e^{\Psi(\chi)})^2(e^{\Phi(1)})^2.$$
(30)

Similarly, the upper-right panel shows the merger rate along the $\cos\theta_1 = \cos\theta_2 = \cos\theta$ diagonal at fixed spin magnitudes ($\chi_1 = \chi_2 = 0.1$) and the same reference masses and redshift:

$$\mathcal{R}(\cos\theta_1, \cos\theta_2 = \cos\theta) = r(e^{\Psi(0.1)})^2(e^{\Phi(\cos\theta)})^2.$$
(31)

We choose to plot results along the $\chi_1 = \chi_2$ and $\cos\theta_1 = \cos\theta_2$ diagonals to mitigate systematic modeling uncertainties, in much the same way that we plot merger rates conditioned on specific values of other parameters rather than marginalizing over them. The rate of black hole mergers as a function of χ_1 only (marginalized over χ_2), for instance, is strongly affected by assumptions regarding spin pairing, which tend to differ widely across the literature. For better comparison with other work, however, we also show in the lower row the implied probability distributions on individual component spin magnitudes and tilts. Since we assume that component spins are independently and identically distributed, the spin magnitude and tilt probability distributions are given by

$$p(\chi) = \frac{e^{\Psi(\chi)}}{\int_0^1 e^{\Psi(\chi')}d\chi'}$$
(32)

and

$$p(\cos\theta) = \frac{e^{\Phi(\cos\theta)}}{\int_{-1}^1 e^{\Phi(\cos\theta')}d\cos\theta'}.$$
(33)

Note that Eqs. (30) and (31) are proportional to the *squares* of Eqs. (32) and (33), respectively.

From Fig. 8, we can make the following parameter-free statements regarding the binary black hole spin distribution:

(1) *The binary black hole merger rate is maximized at low spin magnitudes.* As in Sec. III, we can evaluate the robustness of this statement by comparing mean merger rates in different intervals. We find, for example, that our inferred rate of mergers with $0 \leq \chi \leq 0.2$ is greater than the rate of mergers across $0.6 \leq \chi \leq 0.8$ for each of our 4500 posterior samples on $\mathcal{R}(\chi)$. In Fig. 9, we additionally show the ensemble of cumulative distribution functions corresponding to our posterior on $p(\chi)$ from Fig. 8. We find the 50th percentile to occur at $\chi_{50\%} = 0.21^{+0.07}_{-0.07}$, such that half of black holes have spin magnitudes below $\chi \lesssim 0.2$. The $p(\chi)$ distribution shown in Fig. 8 furthermore suggests that the spin magnitude distribution may actually peak near $\chi \approx 0.2$; the recovered mean (shown in black) increases slightly in this region, and the upper bound on $p(\chi)$ is elevated in the range $0.2 \lesssim \chi \lesssim 0.25$. Neither of these features is statistically significant though; only 66% of traces give larger integrated probability in the $0.15 \lesssim \chi \lesssim 0.35$ interval than in the $0 \lesssim \chi \lesssim 0.15$ interval. Thus, the spin magnitude distribution is consistent with a peak global maximum at $\chi \approx 0$.

(2) *No special features required at $\chi = 0$ or $\chi = 1$.* Although binary black holes exhibit a preference for small spins, the data do not require sharp or discontinuous excesses of nonspinning or maximally spinning black holes. The possible existence of these features has been the subject of much scrutiny. Initial work found that gravitational-wave data were consistent with two distinct subpopulations: a "spike" comprising the majority of the binary population and a secondary broad subpopulation centered at $\chi \approx 0.5$ and possibly extending to large spins [56]. Later work further asserted that such features were in fact *required* by the data [57,58]. In addition, both Refs. [56,57] suggested that the failure by other analyses to properly model a zero-spin subpopulation led to spurious identification of spin-orbit misalignment among the black hole population (to be discussed further below). Follow-up investigations, however, have since concluded that the data remain agnostic about zero-spin or rapidly spinning subpopulations [6,15,16,60].

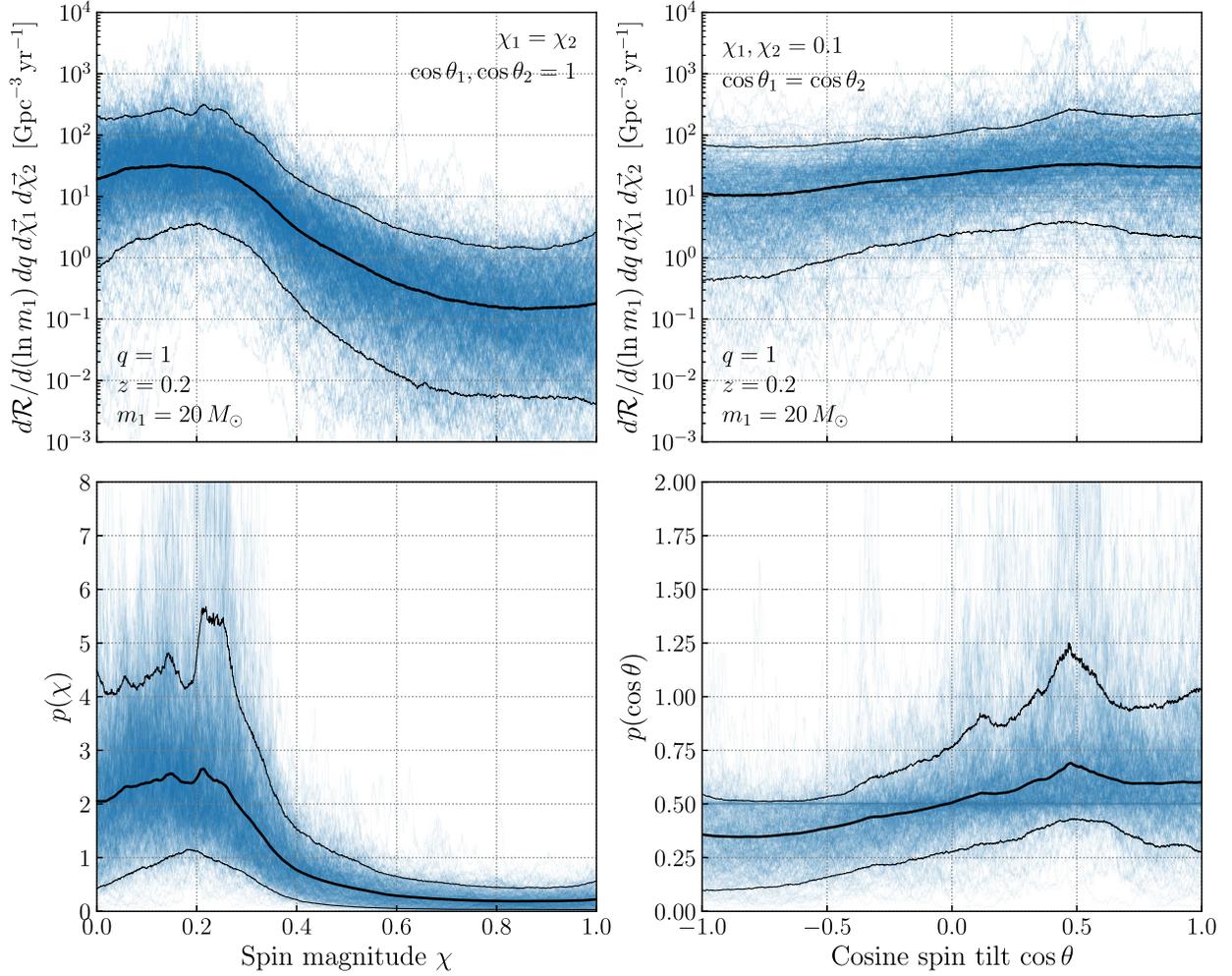In our Fig. 8, we see no indication of an excess of nonspinning systems, nor do we see

FIG. 8.   Top panels: merger rate of binary black holes as a function of component spin magnitudes (left) and spin-orbit misalignment angles (right), as inferred using our autoregressive model defined in Eq. (29). In the axes labels, we use the shorthand $d\vec{\chi}_1 \equiv d\chi_1 d\cos\theta_1$ to indicate a density over both spin magnitude and cosine tilt. Specifically, the rates shown are that of binaries with equal component spin magnitudes [$\chi_1 = \chi_2 = \chi$; see Eq. (30)] or tilts [$\cos\theta_1 = \cos\theta_2 = \cos\theta$; Eq. (31)], each evaluated at fixed reference masses and redshift ($m_1 = 20 M_\odot$, $q = 1$, and $z = 0.2$). The bottom panels show the corresponding probability distributions on component spin magnitudes and tilts among black hole binaries. Within each panel, the central black curve marks the mean inferred rate or probability, while outer black curves bound 90% credible intervals. We see that spin magnitudes are well described by a unimodal distribution that peaks at low values, with no sign of an excess of nonspinning ($\chi = 0$) or near-maximally spinning black holes. Meanwhile, the rate of binary mergers is nonzero across the full range of misalignment angles, with a spin-tilt distribution that is possibly (but not necessarily) isotropic. While there also appears to be a possible excess of black holes with $\cos\theta \approx 0.4$, this feature is not statistically significant.

any feature suggesting a subpopulation of rapidly spinning black holes. There may *exist* a small number of rapidly spinning black holes; as illustrated in Fig. 9, we infer the 95th percentile of the spin magnitude distribution to occur at $\chi_{95\%} = 0.72^{+0.17}_{-0.25}$. We emphasize, however, that there is no observational evidence that these systems comprise a physically distinct subpopulation and not simply an extended tail of a single, predominantly low-spin population. Consistent results have also been found when alternatively using splines to flexibly model the black hole spin distribution [20,21].

Although an excess of zero-spin systems is not *ruled out*, the current lack of discernible features at $\chi = 0$ or $\chi = 1$ is in possible tension with common assumptions in the population synthesis of compact binaries [61–63]: that efficient angular-momentum transport yields isolated black holes born with very small (e.g., $\chi \lesssim 0.1$) [64,65] or vanishing ($\chi \lesssim 0.01$) natal spin magnitudes [66]. Very small natal spin magnitudes should yield a sharp excess of low- or nonspinning systems in the binary black hole spin distribution. Meanwhile, if some fraction of mergers arise from isolated stellar binaries, then late-time tidal spin-up of the second-born black hole's
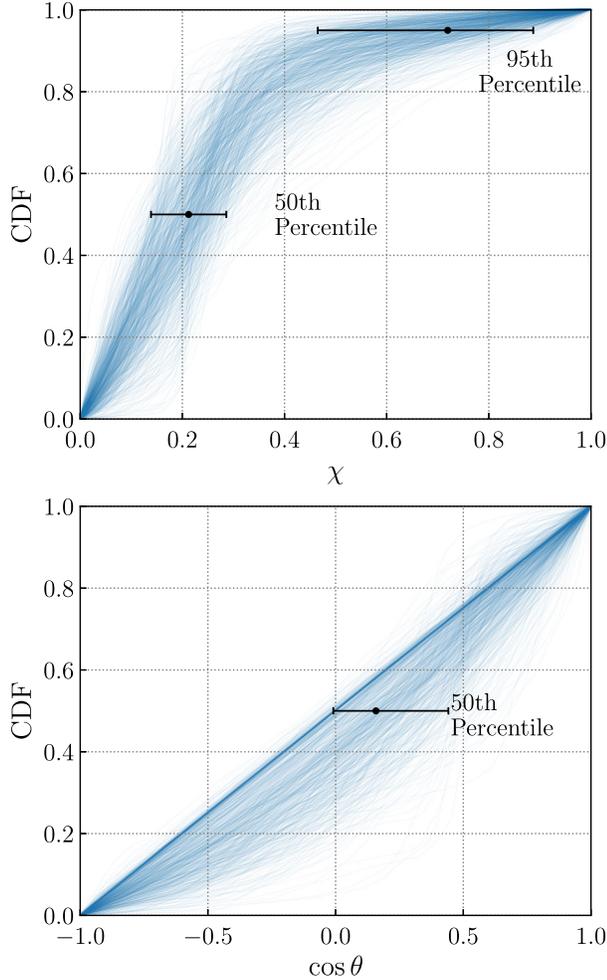
FIG. 9. Cumulative distribution functions of binary black hole spin magnitudes (top panel) and cosine tilt angles (bottom panel), corresponding to the probability distributions shown in Fig. 11. For reference, we mark median estimates of the 50th and 95th percentiles in the spin magnitude distribution, occurring at $\chi_{50\%} = 0.21^{+0.07}_{-0.07}$ and $\chi_{95\%} = 0.72^{+0.17}_{-0.25}$, respectively. We also indicate the measured 50th percentile of the $\cos\theta$ distribution, occurring at $\cos\theta_{50\%} = 0.16^{+0.28}_{-0.17}$, with $\cos\theta_{50\%} > 0$ at 88% credibility.

progenitor can override otherwise efficient angular momentum loss, yielding a secondary subpopulation of black holes with spins up to $\chi \approx 1$ [61,62,67,68]. The absence of such features in current data may suggest that angular momentum transport is less efficient than usually expected.

(3) *The merger rate is nonzero at $\chi = 0$.* Despite the fact that there is no *excess* of systems with vanishing spin, the binary black hole merger rate is confidently nonzero at $\chi = 0$. This finding is in conflict with commonly used parametric models that assume component spins follow nonsingular Beta distributions [6,9,57,69], which, by definition, require that $p(\chi) = 0$ at $\chi = 0$; see Fig. 11 and further

discussion below. [70] The fact that the spin magnitude is nonzero at $\chi = 0$ may have implications for the processes by which black holes acquire their spins. If black holes acquire their spins via stochastic or incoherent isotropic processes (e.g., random bombardment by gravity waves soon before core collapse [71,72] or statistically isotropic fallback accretion), then the spin magnitude distribution should have a Maxwellian-like form $p(\chi) \propto \chi^2$ near $\chi = 0$. The fact that this is not seen suggests, instead, that black hole spins originate from longer-lived or directionally coherent processes [73,74].

(4) *Black holes exhibit a broad range of spin-orbit misalignment angles.* As illustrated in the upper- and lower-right panels of Fig. 8, we infer a nonzero merger rate across the full range of $\cos\theta$. Using our autoregressive constraints on $\mathcal{R}(\cos\theta)$, we estimate that $41^{+9}_{-17}\%$ of black hole spins are misaligned by more than 90° with respect to binaries' orbital angular momenta and that the rate of mergers with at least one component spin tilted by $\theta > 90°$ is $17.0^{+10.7}_{-7.4}$ Gpc^{-3} yr^{-1}. Past studies using both strongly parametrized models [6,9,15,75] and flexible splines [20,21] have also concluded that the binary black hole population exhibits significant spin-orbit misalignment. The results presented here, obtained under our highly agnostic and parameter-free autoregressive model, corroborate these conclusions.

(5) *A perfectly isotropic distribution is moderately disfavored.* As seen in Fig. 8, both the merger rate $\mathcal{R}(\cos\theta)$ and the probability distribution $p(\cos\theta)$ have a tendency to increase towards positive $\cos\theta$. In Fig. 9, we show the corresponding cumulative distribution of $\cos\theta$ and the inferred median $\cos\theta$ among the black hole population. We find this median to be $\cos\theta_{50\%} = 0.16^{+0.28}_{-0.17}$, with $\cos\theta_{50\%} > 0$ for 88% of our posterior samples (the *mean* value of $\cos\theta$ is also positive at comparable credibility). These results somewhat disfavor a purely isotropic component spin distribution, although isotropy cannot yet be ruled out.

(6) *A possible excess of systems with $\cos\theta \approx 0.4$?* As identified in Ref. [75], we also see a possible excess of black holes with $\cos\theta \approx 0.4$. We find that although this feature is *possible*, it is not required by the data. Following our procedure from Sec. III, we can evaluate the significance of the $\cos\theta \approx 0.4$ peak by asking what fraction of posterior samples give a higher mean probability in a window centered on the peak than in windows at both higher and lower $\cos\theta$ values. As shown in Fig. 10, only 37% of samples are consistent with a peak at $\cos\theta \approx 0.4$. While the probability distribution of spin tilts is very likely to *increase* between $\cos\theta \approx -0.1$ and
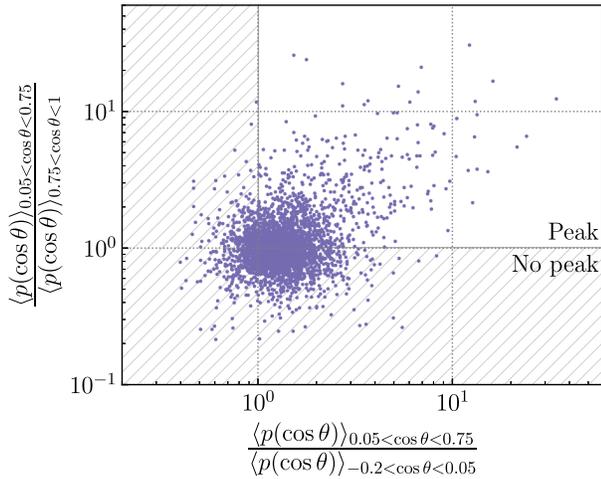
FIG. 10. Evaluation of the significance of the $\cos\theta \approx 0.4$ peak in Fig. 8. For each probability distribution $p(\cos\theta)$ in Fig. 8, we show the ratios between the mean probability in the window $0.05 < \cos\theta < 0.75$ (centered on the possible peak) and the mean probabilities across adjacent windows at smaller and larger $\cos\theta$. When a peak is present, both ratios should be greater than 1, corresponding to the upper-right quadrant. We find that only 37% of posterior samples fall in this quadrant, indicating that the $\cos\theta \approx 0.4$ peak is not statistically significant.

$\cos\theta \approx 0.4$, few samples exhibit the subsequent drop necessary for a peak.

B. Discussion

Figure 11 compares our flexible autoregressive inference with results from the strongly parametrized DEFAULT model [69] presented in Ref. [6]. In this model, component spin magnitudes are independently and identically drawn from a Beta distribution, while spin tilts are drawn from a mixture between isotropic and preferentially aligned subpopulations. The two approaches generally yield similar conclusions, with some notable exceptions. First, as noted above, the DEFAULT model is defined such that $p(\chi)$ is necessarily zero at $\chi = 0$. Our autoregressive results indicate that this is likely not the case; we infer a nonzero rate of mergers with $\chi = 0$ (although no *excess* of such mergers as one might expect if isolated black holes have vanishing natal spins). Second, strongly parametrized approaches typically require the $\cos\theta$ distribution to be either isotropic or peaked at $\cos\theta = 1$. As illustrated in, e.g., the bottom right-hand panel of Fig. 11, the data tell a more complicated story, with a possible (albeit statistically insignificant) feature at intermediate $\cos\theta$ values. See Ref. [75] for further investigations of this feature.

Finally, it is instructive to compare the behavior of $\mathcal{R}(\cos\theta)$, in the top-right panel of Fig. 11, with that of $p(\cos\theta)$, in the bottom-right panel. Studies of the black hole spin distribution sometimes include the following seemingly inconsistent statements: (i) that an isotropic $\cos\theta$ distribution is disfavored but cannot be ruled out and (ii) that our knowledge of $p(\cos\theta)$ is accurately reflected in the red band in the lower-right panel of Fig. 11. Figure 11, though, seems to show, unambiguously, that $p(\cos\theta)$ is an increasing function of $\cos\theta$, in conflict with the first of the two statements above.

The resolution to this apparent paradox involves the fact that we directly measure $\mathcal{R}(\cos\theta)$, not the normalized probability distribution $p(\cos\theta)$. Although spin isotropy is disfavored, it is evident in Fig. 11 that a flat $\mathcal{R}(\cos\theta)$ cannot yet be fully ruled out. The renormalized probability density $p(\cos\theta)$ can inadvertently obscure this fact: Although there may exist many distinct posterior samples that yield isotropic $\mathcal{R}(\cos\theta)$ (e.g., flat traces at different vertical positions within the red or blue bands), each of these possibilities is mapped to the *same* function, $p(\cos\theta) = 1/2$, upon normalization. Thus, hidden behind the "uncertainty bands" in the lower-right panel of Fig. 11 is a very uneven *density* of possibilities, with a high number of individual draws stacked directly on $p(\cos\theta) = 1/2$. Because the uncertainty bounds do not communicate this density, the result is a figure that appears to indicate an unambiguous measurement of anisotropy. In order to avoid this counterintuitive behavior, we recommend that measurements of the $\cos\theta$ distribution be shown as constraints on both the probability density $p(\cos\theta)$ *and* the merger rate $\mathcal{R}(\cos\theta)$.

While our autoregressive model makes minimal physical assumptions, there remain two caveats to consider when interpreting the above results. First, we have chosen to model component spins as independently and identically distributed. This assumption is broken in situations like tidal spin-up of field binaries. We note that our assumption of independence and identicality is purely a choice of model and not a limitation of the method; one could consider, instead, adopting *separate* autoregressive processes for the distributions of primary and secondary spin magnitudes and tilts.

Second, our autoregressive model necessarily imposes a degree of continuity in the merger rate as a function of χ and $\cos\theta$. This continuity could, in principle, obscure very sharp or discontinuous features in the black hole spin distributions. It is therefore reasonable to ask if our conclusions above are being driven by continuity conditions rather than informative data. This question is particularly critical when interpreting our conclusions regarding the lack of sharp features near $\chi \approx 0$; is our nondetection of such features significant, or do they fall outside the coverage of our model? In Appendix C, we conduct a mock data challenge to test the ability of our autoregressive model to recover a sharp excess of non-spinning black holes. Although the resolution of our results is, at times, limited by the processes' finite scale length τ, we find that we can successfully identify narrow excesses
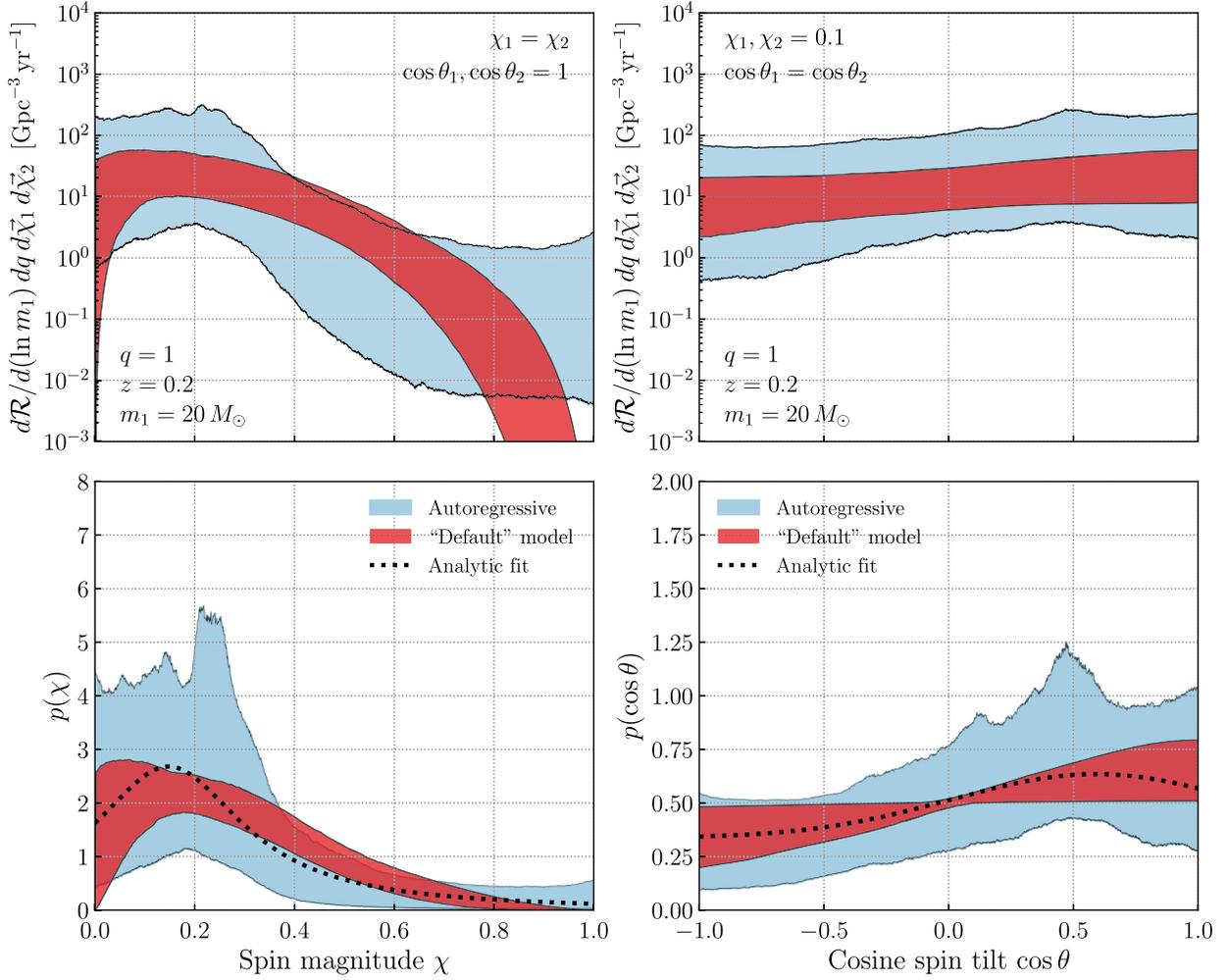
FIG. 11. Comparison of the binary black hole spin distributions inferred using our autoregressive model (blue) and that recovered by a strongly parametrized approach (red, the DEFAULT model of Ref. [6]). As in Fig. 8, the top row shows the binary merger rate as a function of component spin magnitude and spin-orbit tilt angle, at fixed m_1, q, and z, while the lower row shows the corresponding probability distributions. Component spins are assumed to be independently and identically distributed. Overall, there is good reasonable qualitative agreement between both sets of results; each recovers similar merger rates across the range of $\cos\theta$ values and for $0.1 \lesssim \chi \lesssim 0.3$. At the same time, the autoregressive results indicate that the merger rate remains finite for both smaller and larger spin magnitudes, whereas the parametric model requires *a priori* that it vanishes as $\chi \to 0$ and $\chi \to 1$.

or bimodalities in the merger rate arising from a population of nonspinning systems, should it exist.

A related question is the degree to which we can trust extended tails appearing in our autoregressive measurements of the spin-dependent merger rate. As also demonstrated in Appendix C, our autoregressive process never goes completely to zero, as this would correspond to $\ln\mathcal{R} \to -\infty$. Consequently, are the tails in Fig. 8 towards large χ and negative $\cos\theta$ physically meaningful, or do they arise from our prior modeling assumptions? Within Appendix C, we find that, in the absence of observations, the recovered merger rates asymptotically approach a value corresponding to $N_{\exp} \lesssim 1$ total expected detections (integrated across the region of interest). We can leverage this behavior to gauge the extent to which tails in our χ and

$\cos\theta$ distributions are prior or likelihood dominated. Specifically, we use our posteriors on $\mathcal{R}(\chi_i)$ and $\mathcal{R}(\cos\theta_i)$ to compute expected detection rates at large χ and small $\cos\theta$ and identify the threshold spin magnitude and tilts beyond which we expect fewer than $N = 2$ component spins to arise in our sample; these values mark the boundaries beyond which our results are likely prior dominated. This calculation is described in more detail in Appendix E, and it also accounts for the influence of selection effects on the observed distribution of binary parameters.

Our measurements of the spin magnitude distribution imply that we expect $N \leq 2$ detections with at least one component spin magnitude falling above $\chi \geq 0.90^{+0.08}_{-0.26}$, where the uncertainties reflect our uncertain recovery of the

spin magnitude distribution. Thus, our recovered spin magnitude distribution is likely prior dominated at $\chi \gtrsim 0.9$, although prior effects may also become important above $\chi \gtrsim 0.65$ under a more conservative interpretation. Meanwhile, we find that our results, on average, predict fewer than two detections with $\cos\theta \leq -0.96^{+0.12}_{-0.03}$, implying that our posterior on the spin-tilt distribution is likelihood dominated across nearly the full range of $\cos\theta$ values. One might choose more conservative thresholds by instead identifying values beyond which fewer than $N = 4$ detections are predicted; these occur at $\chi \geq 0.82^{+0.13}_{-0.29}$ and $\cos\theta \leq -0.91^{+0.19}_{-0.06}$.

C. Suitable parametric models

When a standard strongly parametrized model is required, we find that our autoregressive measurement of $p(\chi)$ is well fit by a truncated Gaussian or a truncated Lorentzian,

$$p(\chi) = \frac{\mathcal{C}}{\gamma}\left[1 + \left(\frac{\chi - \chi_0}{\gamma}\right)^2\right]^{-1}, \qquad (34)$$

with normalization

$$\mathcal{C} = \left[\tan^{-1}\left(\frac{1 - \chi_0}{\gamma}\right) + \tan^{-1}\left(\frac{\chi_0}{\gamma}\right)\right]^{-1}, \qquad (35)$$

and $p(\cos\theta)$ by a mixture between isotropic and Gaussian components,

$$p(\cos\theta) = \frac{f_{\mathrm{iso}}}{2} + (1 - f_{\mathrm{iso}})N_{[-1,1]}(\cos\theta|\mu, \sigma), \qquad (36)$$

where $N_{[-1,1]}(\cos\theta|\mu, \sigma)$ indicates a truncated Gaussian normalized on the interval $-1 \leq \cos\theta \leq 1$. Equation (36) is the same as the DEFAULT spin-tilt distribution [69] but with a freely varying mean as advocated in Ref. [75]. A least-squares fit of our results to Eqs. (34) and (36) yields best-fit parameters

$$\chi_0 = 0.15,$$
$$\gamma = 0.18,$$
$$f_{\mathrm{iso}} = 0.67,$$
$$\mu = 0.59,$$
$$\sigma = 0.58. \qquad (37)$$

These fits describe marginal probability distributions and are therefore valid at any choice of m_1, q, and z.

VI. STOP FOUR: EFFECTIVE SPINS

Although component spin magnitudes and spin-orbit misalignment angles have a clear physical interpretation,

they are particularly difficult to measure using gravitational waves. Easier to directly measure are various effective spins: derived parameters that, while less physically interpretable, more directly govern a gravitational wave's morphology. These effective parameters include the effective inspiral spin [76,77],

$$\chi_{\mathrm{eff}} = \frac{\chi_1 \cos\theta_1 + q\chi_2 \cos\theta_2}{1 + q}, \qquad (38)$$

and the effective precessing spin [78],

$$\chi_{\mathrm{p}} = \mathrm{Max}\left(\chi_1 \sin\theta_1, \frac{3 + 4q}{4 + 3q}q\chi_2 \sin\theta_2\right). \qquad (39)$$

Here, χ_{eff} quantifies the degree of spin projected parallel to a binary's orbital angular momentum, while χ_{p} approximately quantifies the degree of in-plane spin (and hence more directly controls the degree of spin-orbit precession). Although χ_{eff} and χ_{p} are less manifestly physical than the component spin magnitudes and tilts (much like the relationship between a binary's chirp mass and component masses), they act as signposts by which to identify categorical features of the compact binary spin distribution. Negative χ_{eff}, for example, can arise only if one or both component spins are inclined by more than 90° with respect to their orbit. Nonzero χ_{p}, meanwhile, can manifest only if a system has at least some in-plane spin, such that $\sin\theta > 0$.

Just as we have applied our autoregressive model to nonparametrically infer the component spin magnitude and tilt distributions, we can use our autoregressive model to measure the distribution of these spin parameters. If $\Psi(\chi_{\mathrm{eff}})$ and $\Phi(\chi_{\mathrm{p}})$ are autoregressive functions of χ_{eff} and χ_{p}, respectively, then our merger rate model will be of the form

$$\mathcal{R}(\ln m_1, q, \chi_{\mathrm{eff}}, \chi_p; z)$$
$$= r\frac{f(m_1)p(q)}{f(20M_\odot)}\left(\frac{1 + z}{1 + 0.2}\right)^\kappa [e^{\Psi(\chi_{\mathrm{eff}})}e^{\Phi(\chi_{\mathrm{p}})}], \qquad (40)$$

where we again fall back on parametric models for the dependence of the merger rate on binary masses and redshift. Note that, while we are describing a binary's spin configuration in terms of χ_{eff} and χ_{p}, binary spin is fundamentally six dimensional. Our choice to work in a reduced two-dimensional space requires that we assume some distribution for the remaining 4 degrees of freedom, even if that assumption is implicit. In defining Eq. (40), we indirectly assume that the remaining spin degrees of freedom follow their default parameter estimation priors (uniform spin magnitudes and isotropic directions), conditioned on χ_{eff} and χ_{p}.

A. Features in the black hole effective spin distribution

Figure 12 shows our inference of the χ_{eff} and χ_{p} distributions of binary black holes. As in Fig. 8, the upper row shows the inferred merger rate as a function of χ_{eff} (with fixed χ_{p}; left) and χ_{p} (with fixed χ_{eff}; right). Both rates are evaluated at a fixed reference primary mass, mass ratio, and redshift. The bottom row, meanwhile, shows the corresponding normalized probability distributions of each effective spin parameter. From Fig. 12, we draw the following conclusions:

(1) *The merger rate is nonzero for $\chi_{\text{eff}} < 0$.* Consistent with the results of Sec. V, we find a nonzero merger rate for binaries with $\chi_{\text{eff}} < 0$, suggesting the presence of component spins misaligned by more than 90° with respect to their orbital angular momenta.

We find that $27^{+17}_{-14}\%$ of binary black holes have negative χ_{eff} and that the integrated merger rate of binaries with negative effective spin is $7.7^{+8.2}_{-4.3}$ Gpc^{-3} yr^{-1}. These estimates are comparable to those presented in Ref. [6], which concluded, using a strongly parametrized model, that $29^{+15}_{-13}\%$ of binaries exhibit negative χ_{eff}.

(2) *The χ_{eff} distribution peaks at positive values.* Despite the presence of binaries with negative effective spin, we find that the χ_{eff} distribution is not symmetric about $\chi_{\text{eff}} = 0$ but instead prefers to peak at small but positive values. This preference is significant; among our posterior samples on $\mathcal{R}(\chi_{\text{eff}})$, 98.3% have a larger integrated merger rate in the range $0 \leq \chi_{\text{eff}} \leq 0.1$ than between $-0.1 \leq \chi_{\text{eff}} \leq 0$.
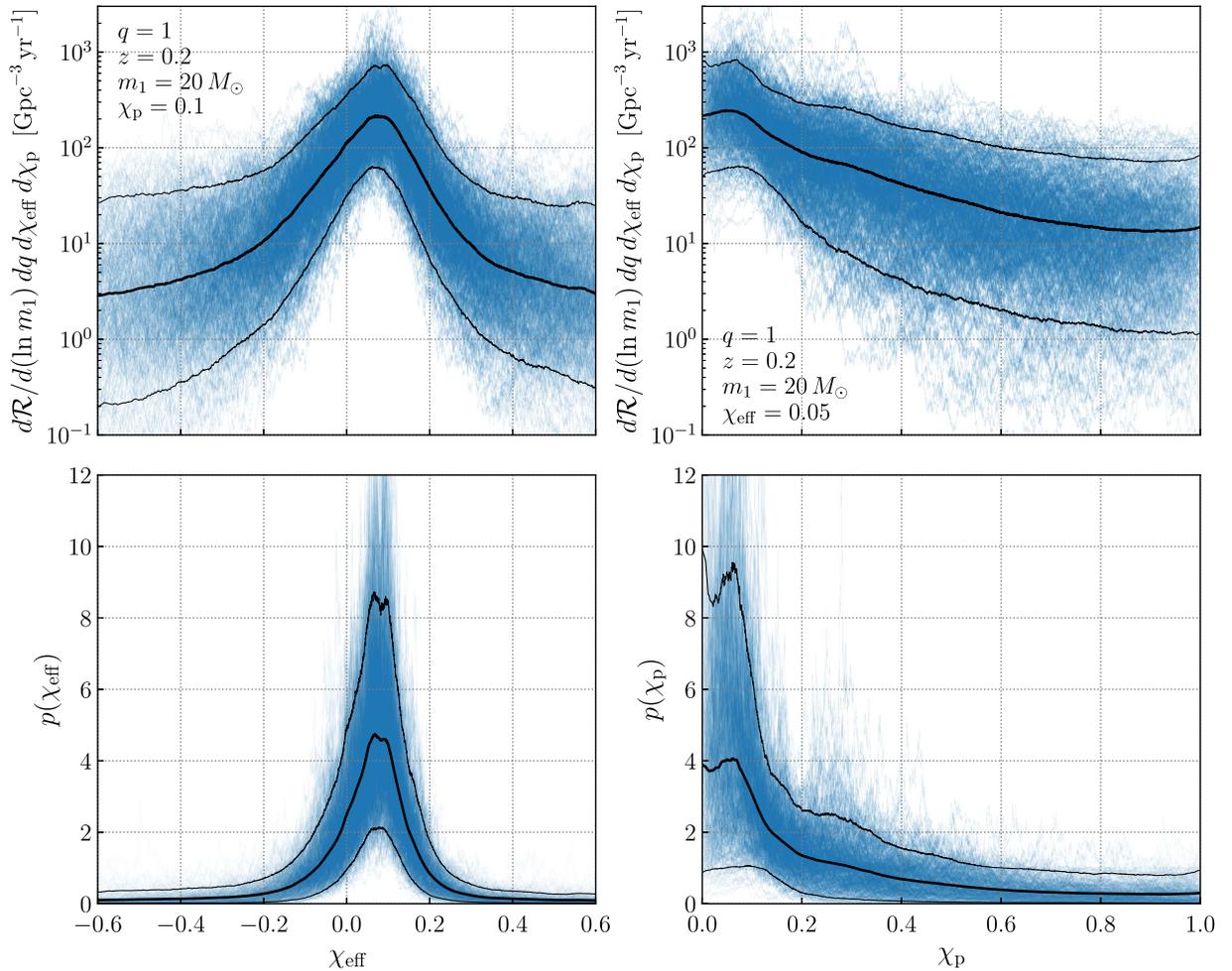

FIG. 12. Top panels: merger rate of binary black holes as a function of effective inspiral spin (χ_{eff}, left) and effective precessing spin (χ_{p}, right), as inferred using our autoregressive model. The rates shown are each evaluated at fixed reference masses and redshift ($m_1 = 20 M_\odot$, $q = 1$, and $z = 0.2$). The bottom panels show the corresponding probability distributions on each effective spin parameter. Within each panel, the central black curve marks the mean inferred rate or probability densities, while outer black curves bound 90% credible intervals. Effective inspiral spins exhibit a unimodal distribution. The center of this distribution prefers to be at positive χ_{eff} but with a nonzero merger rate at $\chi_{\text{eff}} < 0$. The χ_{p} distribution, meanwhile, preferentially peaks toward $\chi_{p} = 0$ but with a shoulder that extends to moderate to large precessing spins.
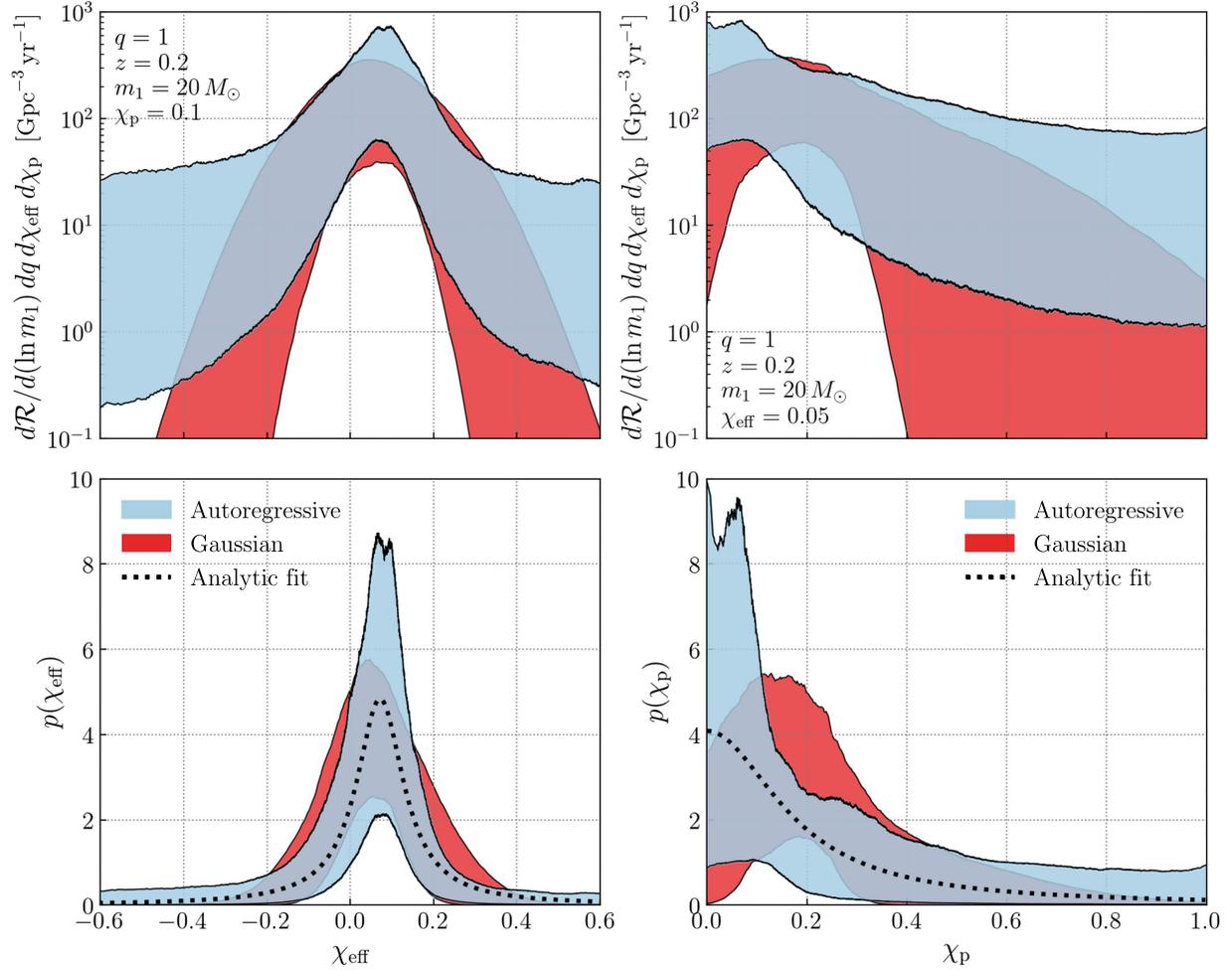
FIG. 13. Comparison of the binary black hole effective spin distributions inferred using our autoregressive model (blue) and that recovered by a strongly parametrized approach (red, the GAUSSIAN spin model of [6]). As in Fig. 12, the top row shows the binary merger rate as a function of effective spin parameters, $\chi_{\rm eff}$ and $\chi_{\rm p}$ (at fixed m_1, q, and z), while the lower row shows the corresponding probability distributions. The merger rates recovered by each approach agree well in the $-0.1 \lesssim \chi_{\rm eff} \lesssim 0.2$ and $0.1 \lesssim \chi_{\rm p} \lesssim 0.3$ ranges, beyond which the GAUSSIAN rates fall to zero much more quickly than our autoregressive inference.

Similarly, the median $\chi_{\rm eff}$ is inferred to be positive for 98.2% of samples [79].

(3) *The binary black hole distribution exhibits nonzero $\chi_{\rm p}$.* Consistent with the measurement of a range of $\cos\theta$ values, we find that the black hole merger rate extends across a wide range of $\chi_{\rm p}$. The percentage of binaries with $\chi_{\rm p} > 0.2$, for example, is $42^{+35}_{-32}\%$.

B. Discussion

We compare our autoregressive measurements to previous strongly parametrized population measurements in Fig. 13. Blue bands show central 90% credible intervals on the rates and probability distributions of $\chi_{\rm eff}$ and $\chi_{\rm p}$ under our autoregressive model, while red bands show results obtained when modeling the $\chi_{\rm eff} - \chi_{\rm p}$ as a bivariate Gaussian in order to measure the mean and standard deviation of each quantity. Our autoregressive $\mathcal{R}(\chi_{\rm eff})$

measurement is, in fact, in reasonable agreement with a Gaussian model, although with extended tails to $\chi_{\rm eff} \gtrsim 0.4$ and $\chi_{\rm eff} \lesssim -0.4$. Both the autoregressive and Gaussian $\chi_{\rm p}$ models, in turn, yield similar merger rates at $\chi_{\rm p} \approx 0.1$, although the Gaussian model appears to vanish too quickly as $\chi_{\rm p} \rightarrow 0$ or 1.

As in the component spin case above, it is valuable to explore whether these extended tails in our $\chi_{\rm eff}$ and $\chi_{\rm p}$ distribution are due to informative data or to the continuity imposed by our autoregressive model. We will once again estimate the regions in which our results are prior dominated by identifying the threshold $\chi_{\rm eff}$ and $\chi_{\rm p}$ values beyond which our posteriors predict fewer than $N = 2$ detections. We find that fewer than two detections are expected at $\chi_{\rm eff} \leq -0.29^{+0.22}_{-0.45}$, at $\chi_{\rm eff} \geq 0.49^{+0.31}_{-0.29}$, and at $\chi_{\rm p} \geq 0.84^{+0.14}_{-0.44}$. This suggests that our results are prior dominated beyond these regions, such that the apparent tension between the
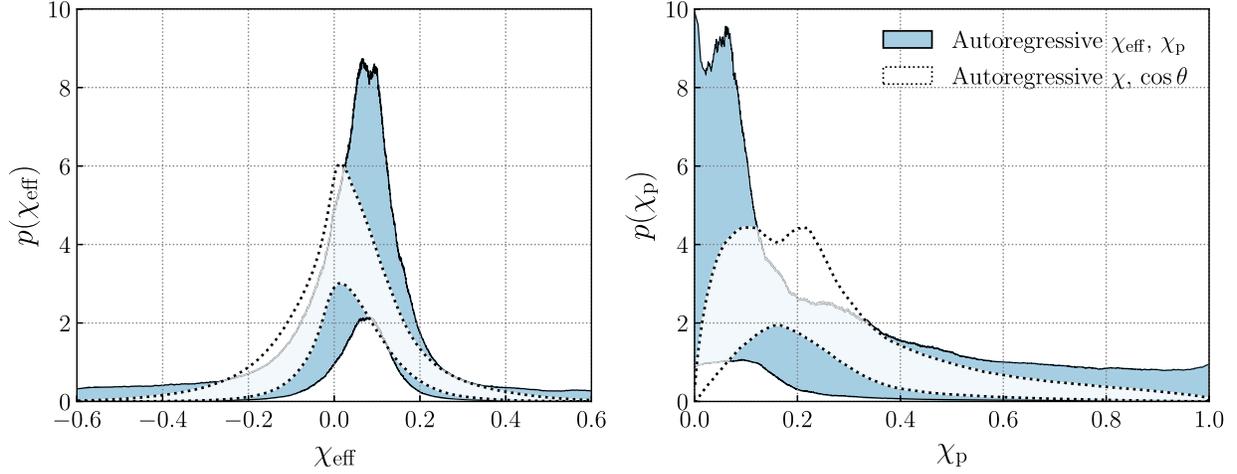
FIG. 14. Comparison between the effective spin distributions obtained through direct autoregressive fits [blue; Eq. (40) in Sec. VI] and the effective spin distributions obtained by instead fitting autoregressive models to the underlying component spin magnitudes and tilts [white; Eq. (29) in Sec. V]. There is not perfect agreement between results. Compared to the direct autoregressive measurement of $p(\chi_{\rm eff})$ and $p(\chi_{\rm p})$, fits to the underlying component spins give an effective inspiral spin distribution that is likely maximized closer to $\chi_{\rm eff} \approx 0$ with a broader tail to positive $\chi_{\rm eff}$, and $p(\chi_{\rm p})$ is meanwhile maximized at larger values of $\chi_{\rm p}$. These differences are expected; each model makes different physical assumptions, and the coordinate transformation from component to effective spins imprints further structure in the derived effective spin distributions [such as the requirement that $p(\chi_{\rm p}) = 0$ at $\chi_{\rm p} = 0$]. At the same time, both sets of results share qualitative features in common, such as the presence of events with negative effective spins. That both results exhibit this feature suggests that it is robust, rather than a spurious result arising from continuity conditions imposed on our autoregressive measurement of $p(\chi_{\rm eff})$.

GAUSSIAN and autoregressive models at very positive and negative $\chi_{\rm eff}$ and very large χ_p is consistent with differing prior assumptions. Furthermore, the fact that we expect $N = 2$ detections with effective spins $\chi_{\rm eff} \lesssim -0.3$ suggests that the inferred presence of negative $\chi_{\rm eff}$ systems is due to informative data, rather than a consequence of prior continuity conditions imposed on our model. However, based on the uncertainties quoted above, we cannot rule out an empirical distribution function that instead only reaches $N = 2$ detections above $\chi_{\rm eff} = -0.09$, such that prior effects set in at $\chi_{\rm eff} \approx -0.1$. In this more conservative approach, is it possible that continuity conditions are "fooling" our model into inferring the presence of negative effective spins where none in fact exist?

If all binary black holes had spin-orbit misalignment angles below $\theta = 90°$ degrees (and hence purely positive effective spins), this could manifest as a sharp truncation in the effective spin distribution at $\chi_{\rm eff} = 0$. In this case, our autoregressive model would indeed struggle to fit such a discontinuity, possibly leading us to incorrectly conclude the existence of systems with negative effective spins. There is, however, a simple remedy to this problem. We should simply move to coordinates in which there are no such discontinuities: component spin magnitudes and tilts. In Sec. V above, autoregressive modeling of the spin-tilt distribution implies that a significant fraction of black holes have misalignment angles greater than 90°, consistent with the need for negative $\chi_{\rm eff}$ identified in this section.

Figure 14 compares our autoregressive measurements of the effective spin distributions in this section (blue) with the effective spin distributions implied by our component spin measurements in Sec. V (white). In order to compute these implied distributions, we again assume that component spins are independently and identically distributed. The white and blue distributions are not identical. This is expected; each measurement corresponds to fundamentally distinct models, and further structure is necessarily imposed by the coordinate transformation from component to effective spins [80]. At the same time, we see the same qualitative features in both sets of results: a $\chi_{\rm eff}$ distribution extending to negative values and a broad $\chi_{\rm p}$ distribution. That these features emerge whether we choose to describe the black hole population via its component spins or effective spins suggests that these conclusions are robust and not due to modeling systematics.

C. Suitable parametric models

When a strongly parametrized model is needed for $p(\chi_{\rm eff})$, we find our autoregressive result to be well approximated by a truncated Gaussian with mean and standard deviation

$$\mu = 0.07,$$
$$\sigma = 0.09, \tag{41}$$

or a truncated Lorentzian [see Eq. (34)] with

$$\chi_0 = 0.07,$$
$$\gamma = 0.07; \tag{42}$$

this latter fit is shown as a dotted line in the lower-right panel of Fig. 13. Similarly, $p(\chi_p)$ can be approximated by either a truncated Gaussian or Lorentzian, with best-fit parameters

$$\mu = 0.0,$$
$$\sigma = 0.23, \tag{43}$$

and

$$\chi_0 = 0.0,$$
$$\gamma = 0.18, \tag{44}$$

respectively, the latter of which is shown in Fig. 13. As in previous sections, these fits are valid at any choice of m_1, q, or z.

VII. CONCLUSION

In this paper, we have developed and demonstrated a novel means of measuring the population properties of merging binary black holes. By describing the black hole merger rate as a stochastic process, we hierarchically inferred the black hole mass, redshift, and spin distributions without resorting to strongly parametrized models that *a priori* assume some particular structure. The advantage of highly flexible models like autoregressive processes is twofold. They allow us to agnostically study the "known unknowns," like theoretically predicted features in the black hole population, but also reveal the "unknown unknowns," unexpected and impactful features that may otherwise be missed by standard strongly parametrized approaches.

We accordingly searched for expected and unexpected features alike in the distributions of binary black hole masses, redshifts, and spins. Our results reiterated known features in the black hole mass spectrum (peaks at approximately 10 and $35M_\odot$) but also revealed a more nuanced structure like an additional steepening of $\mathcal{R}(m_1)$ towards high masses. We found signs of unexpected structure in the redshift distribution of binary black holes, recovering a merger rate that prefers to remain flat at low redshifts followed by steeper growth at $z \gtrsim 0.5$. In addition, our autoregressive results offered a direct and model agnostic look at the black hole spin distribution, revealing features like severe spin-orbit misalignment and a unimodal spin magnitude distribution that have previously been controversial.

A challenge that arises when using flexible models is how exactly to translate results [e.g., our posterior on $\mathcal{R}(m_1)$] into statements about physical features and their significance. We find it useful to conceptually distinguish between two steps: (i) data fitting and (ii) feature extraction. When performing hierarchical inference with strongly parametrized models, these two steps are accomplished simultaneously. A clear example is the POWER LAW+PEAK model for $\mathcal{R}(m_1)$, whose parameters directly encode the location, width, and height of a possible Gaussian peak. Fitting the POWER LAW+PEAK model to data, therefore, automatically extracts information about the feature of interest. When using highly flexible models, on the other hand, data fitting and feature extraction are necessarily distinct. Although hierarchically fitting our autoregressive model yields, for instance, the mass spectrum shown in Fig. 3, this result offers no immediate information about the presence and/or significance of possible features. Instead, we need to visually inspect our results and devise further tests or summary statistics to make any quantitative statements about the features we see. A major focus of our work has accordingly been the use of parameter-free summary statistics, like the ratios of merger rates in adjacent bins, to identify and characterize the features summarized above. These parameter-free techniques for feature extraction can be employed for any model and additionally offer a means of directly comparing results obtained under two or more different models (strongly parametrized or not).

While highly flexible models like ours enable a very agnostic exploration of the compact binary population, we do not necessarily advocate for replacing standard strongly parametrized models. Instead, we envision using both strongly parametrized and flexible models in a cyclic development process: flexible models enable the identification of possible new features, which are followed up and characterized using targeted strongly parametrized models, whose validity is finally rechecked with flexible models as new data become available. In the spirit of this cyclic development, in each section above we have offered refined strongly parametrized models that capture the range of features identified in our autoregressive results.

One limitation of the autoregressive model employed here is the fact that it is fundamentally one dimensional. Although we can simultaneously measure the dependence of the merger rate on different binary parameters, each with its own autoregressive process, this approach cannot capture any intrinsic correlations among parameters. As strongly parametrized models begin to identify possible correlations between binary parameters [17,54], flexible population models that can operate in higher dimensions will be critical in following up on these results and agnostically identifying new correlations. Some alternative approaches, like spline-based [19,20] or binned [22,23] models, can be very easily extended to more than $n = 1$ dimension but likely become computationally infeasible when n becomes large. Future work will involve the exploration of multidimensional stochastic processes as

tools with which to measure the merger rate across the complete higher-dimensional space of binary black hole parameters.

The code used for this study is hosted on GitHub at [81], and data produced by our analyses can be downloaded from Zenodo at [82].

APPENDIX A: MORE ON AUTOREGRESSIVE MODELS

As discussed in the main body, in this work, we agnostically model the rate density of compact binaries as an autoregressive process. Given a merger rate $\ln \mathcal{R}_i$ at some mass $\ln m_i$, Eq. (6) offers a prescription with which to randomly propose a merger rate $\ln \mathcal{R}_{i+1}$ at the next mass $\ln m_{i+1}$ of interest. We still need to initialize this process, though, picking some initial rate $\ln \mathcal{R}_1$ at the smallest mass $\ln m_1$ considered in our sample. This initial value is randomly drawn via

$$\ln \mathcal{R}_1 \sim N(\ln r, \sigma). \tag{A1}$$

The exact form of Eq. (6) and the definitions of c_i and w_i [Eqs. (7) and (8), respectively] are chosen to guarantee that all subsequent rates $\ln \mathcal{R}_i$ have the same marginal prior as $\ln \mathcal{R}_1$, such that the autoregressive process is stationary. For example, from Eq. (6), the prior expectation value of $\ln \mathcal{R}_i$ is given by

$$\langle \ln \mathcal{R}_i \rangle = \ln r + c_i[\langle \ln \mathcal{R}_{i-1} \rangle - \ln r], \tag{A2}$$

where we have used the fact that $\langle w_i \rangle = 0$. From above, though, we know that our initial point satisfies $\langle \ln \mathcal{R}_1 \rangle = \ln r$, implying

$$\langle \ln \mathcal{R}_i \rangle = \ln r \tag{A3}$$

for all i. Similarly, the variance of $\ln \mathcal{R}_i$ is

$$\begin{aligned} \mathrm{Var}(\ln \mathcal{R}_i) &= c_i^2 \mathrm{Var}(\ln \mathcal{R}_{i-1}) + \mathrm{Var}(w_i) \\ &= e^{-2\Delta_i/\tau} \mathrm{Var}(\ln \mathcal{R}_{i-1}) + \sigma^2(1 - e^{-2\Delta_i/\tau}). \end{aligned} \tag{A4}$$

Consider the $i = 2$ case. From Eq. (A1), we know that the variance of $\ln \mathcal{R}_1$ is σ^2, giving

$$\begin{aligned} \mathrm{Var}(\ln \mathcal{R}_2) &= e^{-2\Delta_i/\tau}\sigma^2 + \sigma^2(1 - e^{-2\Delta_i/\tau}) \\ &= \sigma^2. \end{aligned} \tag{A5}$$

By induction,

$$\mathrm{Var}(\ln \mathcal{R}_i) = \sigma^2 \tag{A6}$$

for all subsequent i. We can finally consider the covariance between the rates $\ln \mathcal{R}_{i+n}$ and $\ln \mathcal{R}_i$ at two different locations:

$$\begin{aligned} &\mathrm{Cov}(\ln \mathcal{R}_{i+n}, \ln \mathcal{R}_i) \\ &= \langle \ln \mathcal{R}_{i+n} \ln \mathcal{R}_i \rangle - \langle \ln \mathcal{R}_{i+n} \rangle \langle \ln \mathcal{R}_i \rangle \\ &= \langle [\ln r + c_{i+n}(\ln \mathcal{R}_{i+n-1} - \ln r) + w_{i+n}] \ln \mathcal{R}_i \rangle - (\ln r)^2 \\ &= c_{i+n} \langle \ln \mathcal{R}_{i+n-1} \ln \mathcal{R}_i \rangle - c_{i+n}(\ln r)^2 \\ &= c_{i+n} \mathrm{Cov}(\ln \mathcal{R}_{i+n-1}, \ln \mathcal{R}_i). \end{aligned} \tag{A7}$$

To obtain the third line, we used the definition of our autoregressive process to write $\ln \mathcal{R}_{i+n}$ in terms of $\ln \mathcal{R}_{i+n-1}$. To move to the fourth line, we then used the facts that w_{i+n} and $\ln \mathcal{R}_i$ are uncorrelated, that $\langle w_{i+n} \rangle = 0$, and that $\langle \ln \mathcal{R}_i \rangle = \ln r$. Continuing to iterate in this fashion gives

$$\begin{aligned} &\mathrm{Cov}(\ln \mathcal{R}_{i+n}, \ln \mathcal{R}_i) \\ &= \left(\prod_{j=i+1}^{i+n} c_j \right) \mathrm{Cov}(\ln \mathcal{R}_i, \ln \mathcal{R}_i) \\ &= e^{-(\Delta_{i+n} + \Delta_{i+n-1} + \cdots + \Delta_{i+1})/\tau} \mathrm{Var}(\ln \mathcal{R}_i) \\ &= \sigma^2 e^{-(\ln m_{i+n} - \ln m_i)/\tau}. \end{aligned} \tag{A8}$$

Thus, τ is indeed the scale over which the autoregressive process retains significant autocorrelation.

Implementing our hierarchical likelihood model in numpyro [39,40] and jax [38] necessitates efficient proposals of new autoregressive processes drawn from our

prior. To make our discussion more concrete, consider the proposal of an autoregressive process Ψ over (log) black hole masses $\ln m$. As a preprocessing step, define $[\![\ln m]\!]$ to be the sorted union of all of our posterior samples and injections,

$$[\![\ln m]\!] = \mathrm{Sort}(\{\ln m_1\} \cup \ldots \cup \{\ln m_{N_{\mathrm{obs}}}\} \cup \{\ln m_{\mathrm{inj}}\}), \tag{A9}$$

where $\{\ln m_i\}$ is the set of posterior samples associated with event i and $\{\ln m_{\mathrm{inj}}\}$ is the set of masses corresponding to successfully recovered injections. Define n_{pts} to be the length of $[\![\ln m]\!]$. In addition, precompute the set of differences $[\![\Delta]\!]$ between each adjacent pair in $[\![\ln m]\!]$,

$$[\![\Delta]\!] = \mathrm{Diff}\,[\![\ln m]\!], \tag{A10}$$

where $[\![\Delta]\!]$ is of length $n_{\mathrm{pts}} - 1$. Given priors $p(\sigma)$, $p(\tau)$, and $p(\ln r)$ on the standard deviation, length scale, and mean of the autoregressive process Ψ, realizations of Ψ can then be proposed as follows.

(1) Draw $\sigma \sim p(\sigma)$, $\tau \sim p(\tau)$, and $\ln r \sim p(\ln r)$ from their respective priors.
(2) Draw a set $[\![n_i]\!]_{i=1}^{n_{\mathrm{pts}}} \sim N(0,1)$ of values from a unit normal distribution.
(3) Initialize the autoregressive process by defining $\tilde{\Psi}_1 = \sigma n_1$.
(4) Compute sets $[\![w_i]\!] = [\![\sigma(1 - e^{-2\Delta_i/\tau})n_i]\!]$ and $[\![c_i]\!] = [\![e^{-\Delta_i/\tau}]\!]$ for $i \in (2, n_{\mathrm{pts}})$.
(5) Compute all $\tilde{\Psi}_{i>1}$ via iterating $\tilde{\Psi}_i = c_i \tilde{\Psi}_{i-1} + w_i$.
(6) Finally, apply the mean: $[\![\Psi_i]\!] = [\![\tilde{\Psi}_i + \ln r]\!]$.

To enable efficient sampling, the $[\![\Psi_i]\!]$, $[\![w_i]\!]$, and $[\![c_i]\!]$ are generated following noncentered approaches; we directly sample in $[\![\tilde{\Psi}_i]\!]$ and $[\![n_i]\!]$ and then transform to the actual parameters of interest. Once the complete set $[\![\Psi_i]\!]$ of log-merger rates is generated, the sorting performed to obtain Eq. (A9) can be reversed to repartition $[\![\Psi_i]\!]$ back into the merger rates across individual events' posterior samples and found injections.

In some cases, the merger rate $\ln \mathcal{R}(\theta)$ is not well measured at the lowest θ in our set of samples but at some intermediate value. The merger rate as a function of mass, for example, is much better constrained near $m \approx 25 M_\odot$ than at the very lowest masses $m \lesssim 5 M_\odot$. In this case, sampling efficiency is maximized by not initializing our autoregressive process at its leftmost point (as in Step 3 above) but instead initializing the process in the *middle* of our parameter range, near the best-measured rate. In this case, Steps 4–6 above are just repeated twice, once to generate forward steps to the right of our reference point and once to generate backward steps to the left of the reference point.

APPENDIX B: FURTHER INFERENCE DETAILS AND PRIOR CONSTRAINTS

In this appendix, we give additional information about the exact data used in this paper and further details regarding our implementation and inference of the autoregressive population model.

In our analyses, we include binary black holes in the GWTC-3 catalog [1] detected with false alarm rates below 1 yr^{-1}. GWTC-3 contains two events, GW190814 [90] and GW190917, that are likely binary black holes but are known to be outliers with respect to the bulk binary population [6]; we exclude these two events, leaving 69 binary black holes to be included in our analysis. We use publicly available parameter estimation samples hosted by the Gravitational-Wave Open Science Center [91–93] and/or Zenodo. Each binary typically has several distinct sets of associated posterior samples. For events first identified in GWTC-1 [94], we use the "Overall_posterior" samples [95]. For events announced in GWTC-2 [96], we use the "PrecessingSpinIMRHM" samples [97], and for new events in GWTC-3 [1], we use the "C01:Mixed" samples [98]. Each of these sets corresponds to a union of parameter estimation samples from various waveform families. All waveforms include the physical effects of spin precession, although parameter estimation accounting for higher-order modes is only available for GWTC-2 and GWTC-3.

Our hierarchical inference relies on the use of injected signals to characterize search selection effects; see Eq. (12). We use the injection set discussed in Ref. [6,99] characterizing injections as "found" if they are recovered with false-alarm rates below 1 yr^{-1} in at least one search pipeline. Note that the subset of injections performed for the O1 and O2 observing runs does not have associated false-alarm rates only network signal-to-noise ratios ρ. For these events, we consider them "found" if $\rho \geq 10$.

Ensuring convergence of our inference sometimes requires careful regularization of σ and τ. In particular, when performing inference with an autoregressive prior, we find that a common instability is a runaway towards $\sigma \to \infty$ and $\tau \to 0$. The cause of this behaviour can be seen from the hierarchical likelihood defined in Eqs. (9) and (12). In particular, the likelihood is maximized if the merger rates at each posterior sample can grow to $\mathcal{R}(\lambda_{I,j}) \to \infty$ while sending the merger rates at each injection to $\mathcal{R}(\lambda_{\mathrm{inj},i}) \to 0$ (in turn, sending $N_{\mathrm{exp}} \to 0$). This situation is sketched in Fig. 15, in which posterior samples are denoted as filled circles while injections are marked with stars. Strongly parametrized models usually impose strict constraints on the continuity and smoothness of the merger rate, preventing this behavior. Our autoregressive model and similar approaches, in contrast, are explicitly designed to allow for rapid variations in the merger rate, so they are subject to this instability. In particular, our autoregressive inference is
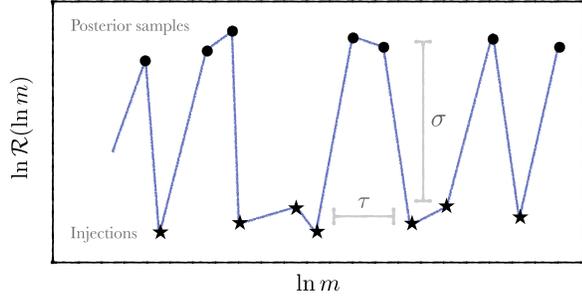
FIG. 15. Illustration of the runaway situation described in Appendix B. When performing inference with an auto-regressive process population model, one can encounter a runaway instability occurring at small scale lengths τ and large variances σ^2. In this regime, the likelihood [Eq. (9)] can be made arbitrarily large by allowing the merger rate at the locations of posterior samples (denoted by circles) to approach $\ln \mathcal{R} \to \infty$ while sending the merger rate at the locations of found injections (stars) to $\ln \mathcal{R} \to -\infty$. This behavior can be combated with additional regularization that penalizes very large ratios $\sigma/\sqrt{\tau}$, as in Eq. (B24).

most susceptible to runaway oscillatory behavior when the processes' variance becomes large and the autocorrelation length becomes small.

This instability can be regulated by placing suitable priors on the parameters governing the autoregressive process. To motivate physically meaningful priors, it is useful to think about how an autoregressive process is allowed to vary between two points. Consider an autoregressive process $\Psi(x)$ with zero mean, variance σ^2, and autocorrelation length τ defined across some parameter x. In addition, let $\delta\Psi = \Psi_2 - \Psi_1$ be the *difference* in the process' values between two points x_1 and x_2 (separated by $\delta x = x_2 - x_1$). The expectation value of $\delta\Psi$ is

$$
\begin{aligned}
\langle \delta\Psi \rangle &= \langle \Psi_2 \rangle - \langle \Psi_1 \rangle \\
&= 0
\end{aligned}
\tag{B1}
$$

since, by definition, the process has zero mean. The expectation value of $\delta\Psi^2$, though, is nonzero:

$$
\begin{aligned}
\langle \delta\Psi^2 \rangle &= \langle (\Psi_2 - \Psi_1)^2 \rangle \\
&= \langle \Psi_2^2 \rangle + \langle \Psi_1^2 \rangle - 2\text{Cov}(\Psi_2, \Psi_1) \\
&= 2\sigma^2 - 2\sigma^2 e^{-\delta x/\tau} \\
&= 2\sigma^2 (1 - e^{-\delta x/\tau}),
\end{aligned}
\tag{B2}
$$

using Eqs. (A6) and (A7) for the variance and covariance of an autoregressive process. It is helpful to consider Eq. (B2) in two different limits. In the limit $\delta x \ll \tau$,

$$
\lim_{\delta x \ll \tau} \langle \delta\Psi^2 \rangle = 2\delta x \frac{\sigma^2}{\tau}.
\tag{B3}
$$

In the opposite limit,

$$
\lim_{\delta x \gg \tau} \langle \delta\Psi^2 \rangle = 2\sigma^2.
\tag{B4}
$$

More generally, we can show that in each of the above limits, $\delta\Psi^2$ is χ^2 distributed with $k = 1$ degrees of freedom. Recall that Ψ_1 and Ψ_2 are related by

$$
\Psi_2 = e^{-\delta x/\tau}\Psi_1 + \sigma(1 - e^{-2\delta x/\tau})^{1/2}n,
\tag{B5}
$$

where $n \sim N(0, 1)$ is drawn from a unit normal distribution. Then,

$$
\delta\Psi^2 = [(e^{-\delta x/\tau} - 1)\Psi_1 + \sigma(1 - e^{-2\delta x/\tau})^{1/2}n]^2.
\tag{B6}
$$

First, consider the $\delta x \ll \tau$ limit. Expanding to lowest order in $\delta x/\tau$,

$$
\begin{aligned}
\delta\Psi^2 &\approx \left(-\frac{\delta x}{\tau}\Psi_1 + \sigma n\sqrt{\frac{2\delta x}{\tau}} \right)^2 \\
&\approx \sigma^2 n^2 \left(\frac{2\delta x}{\tau} \right),
\end{aligned}
\tag{B7}
$$

where the first term is subdominant to the second in the limit of small $\delta x/\tau$. We therefore have

$$
\frac{\delta\Psi^2/\delta x}{2\sigma^2/\tau} \approx n^2
\tag{B8}
$$

such that, by definition, this quantity is chi-squared distributed with 1 degree of freedom:

$$
\frac{\delta\Psi^2/\delta x}{2\sigma^2/\tau} \sim \chi^2(1).
\tag{B9}
$$

Its corresponding expectation value is

$$
\left\langle \frac{\delta\Psi^2/\delta x}{2\sigma^2/\tau} \right\rangle = \frac{\langle \delta\Psi^2 \rangle/\delta x}{2\sigma^2/\tau} = 1;
\tag{B10}
$$

compare to Eq. (B3) above. In the opposite limit, where $\delta x \gg \tau$, Eq. (B6) becomes

$$
\delta\Psi^2 \approx (-\Psi_1 + \sigma n)^2.
\tag{B11}
$$

Note that Ψ_1 is itself drawn from a normal distribution (see Appendix A). We can therefore write

$$
\delta\Psi^2 \approx \sigma^2(n - m)^2,
\tag{B12}
$$

where $m \sim N(0, 1)$ is another normally distributed random variable. The difference $n - m$ is itself a

Gaussian random variable with zero mean and standard deviation $\sqrt{2}$, so

$$\delta\Psi^2 = 2\sigma^2\bar{n}^2, \tag{B13}$$

where $\bar{n} = (n - m)/\sqrt{2} \sim N(0, 1)$ now follows a unit normal distribution. Hence,

$$\frac{\delta\Psi^2}{2\sigma^2} \sim \chi^2(1) \tag{B14}$$

is also chi-squared distributed with 1 degree of freedom and with mean

$$\frac{\langle\delta\Psi^2\rangle}{2\sigma^2} = 1; \tag{B15}$$

compare to Eq. (B4). All together,

$$\begin{cases} \frac{\delta\Psi^2/\delta x}{2\sigma^2/\tau} \sim \chi^2(1) & (\delta x \ll \tau) \\ \frac{\delta\Psi^2}{2\sigma^2} \sim \chi^2(1) & (\delta x \gg \tau). \end{cases} \tag{B16}$$

We use Eq. (B16) to motivate physical priors on σ^2 and τ. First, we might expect the log merger rate to vary by no more than $\delta\Psi_{\max}$ over the full parameter space. Let Δx be the full extent of the parameter space and assume that $\Delta x \gtrsim \tau$, such that we are in the second case in Eq. (B16). Then, our expectation is that

$$q < \frac{\delta\Psi_{\max}^2}{2\sigma^2} \tag{B17}$$

for a chi-squared distributed random variable $q \sim \chi^2(1)$. In particular, we might assert that $\delta\Psi_{\max}^2/2\sigma^2$ occurs at the 99th percentile q_{99} of the chi-squared distribution. The cumulative distribution of a $\chi^2(1)$ distribution is the regularized gamma function $P(1/2; q/2)$, so q_{99} occurs at

$$q_{99} = P^{-1}(1/2; 0.99) \approx 3.32. \tag{B18}$$

Inserting into Eq. (B17) and solving for σ, we have

$$\sigma < \frac{\delta\Psi_{\max}}{\sqrt{2q_{99}}}. \tag{B19}$$

We therefore choose a prior on σ to enforce Eq. (B19). Specifically, we adopt a half-Gaussian prior

$$p(\sigma|\Sigma_\sigma) \propto \exp\left(-\frac{\sigma^2}{2\Sigma_\sigma^2}\right) \tag{B20}$$

in the range $0 \leq \sigma < \infty$. The scale Σ_σ^2 of this prior is chosen so that 95% of our prior weight occurs below the threshold set by Eq. (B19). Specifically,

$$\Sigma_\sigma = \frac{\delta\Psi_{\max}}{2\sqrt{q_{99}}\mathrm{Erf}^{-1}(0.95)}. \tag{B21}$$

Next, consider how we expect our autoregressive process to vary on small scales. In particular, let $\delta\Psi_{\mathrm{event}}$ be the maximum variation we expect in the log rate on the typical interevent distance scale $\delta x \approx \Delta x/N$, where N is the number of events in our catalog. This distance scale is likely smaller than the autoregressive process' correlation length τ, so we now use the first inequality in Eq. (B16), demanding that

$$q < \frac{N\delta\Psi_{\mathrm{event}}^2/\Delta x}{2\sigma^2/\tau}, \tag{B22}$$

for $q \sim \chi^2(1)$. We proceed as above, using Eq. (B22) to define the 99th percentile of q and rearranging to obtain the limit

$$\frac{\sigma}{\sqrt{\tau}} < \delta\Psi_{\mathrm{event}}\sqrt{\frac{N}{2q_{99}\Delta x}}. \tag{B23}$$

We impose this limit by adopting another half-Gaussian prior on the ratio $\sigma/\sqrt{\tau}$,

$$p(\sigma/\sqrt{\tau}|\Sigma_r) \propto \exp\left(-\frac{(\sigma/\sqrt{\tau})^2}{2\Sigma_r^2}\right), \tag{B24}$$

with Σ_r chosen such that 95% of the prior weight occurs below the limit set by Eq. (B23):

$$\Sigma_r = \frac{\delta\Psi_{\mathrm{event}}}{2\mathrm{Erf}^{-1}(0.95)}\sqrt{\frac{N}{q_{99}\Delta x}}. \tag{B25}$$

We can also place limits on the expected length scale τ of our autoregressive processes. We generally expect the merger rate to be smoothly varying across the parameter space of interest, but we also want a prior that will nevertheless allow for rapid, small-scale variations should they be demanded by the data. We accordingly place an unbounded Gaussian prior on $\ln\tau$:

$$p(\ln\tau|\Sigma_{\ln\tau}) \propto \exp\left(-\frac{(\ln\tau - \ln(\Delta x/2))^2}{2\Sigma_{\ln\tau}^2}\right). \tag{B26}$$

This prior is centered at $\ln(\Delta x/2)$, and we set $\Sigma_{\ln\tau}$ by considering the minimum length scale that can be meaningfully constrained by N detections. In particular, the data contain no information about features on scales smaller than the minimum spacing between events. If we consider randomly placing N events across an interval of width Δx, the spacing δx between events will be exponentially distributed:

$$p(\delta x) = \frac{(N/\Delta x)e^{-N\delta x/\Delta x}}{1 - e^{-N}}, \tag{B27}$$

TABLE I. Priors governing the various autoregressive process models used in this paper. For each physical parameter, we give our choices for the scale values characterizing the priors defined in Appendix B: the domain width δx, the maximum variation $\delta\Psi_{max}$ in the log merger rate, the maximum interevent variation $\delta\Psi_{event}$ in the log rate, and the number N of measurements considered. We specifically give $e^{\delta\Psi_{max}}$ and $e^{\delta\Psi_{event}}$, so the quantities listed in the table are directly interpretable as merger rate variations. Also note that while the majority of priors use $N = 69$ (the number of events in our sample), we take $N = 138$ when analyzing the component spin magnitude and tilt distributions, as each binary contributes two component spins to our sample. Given these parameter choices, we also show the derived quantities directly appearing in Eqs. (B20), (B24), and (B26).

Parameter	Section	Δx	$e^{\delta\Psi_{max}}$	$e^{\delta\Psi_{event}}$	N	Σ_σ	$\ln(\Delta x/2)$	$\Sigma_{\ln\tau}$	Σ_r	δx_{min}
$\ln m_1$	Sec. III	4	10^2	2	69	0.91	0.69	4.72	0.57	8.5×10^{-4}
q	Sec. III	1	10^2	2	69	0.91	-0.69	4.72	1.14	2.1×10^{-4}
z	Sec. IV	2	10^2	2	69	0.91	0.0	4.72	0.81	4.2×10^{-4}
χ_1, χ_2	Sec. V	1	10^2	1.5	138	0.91	0.69	5.57	0.94	5.3×10^{-5}
$\cos\theta_1, \cos\theta_2$	Sec. V	2	10^2	1.5	138	0.91	0.0	5.57	0.67	1.1×10^{-4}
χ_{eff}	Sec. VI	2	10^2	2	69	0.91	0.0	4.72	0.81	4.2×10^{-4}
χ_p	Sec. VI	1	10^2	2	69	0.91	-0.69	4.72	1.14	2.1×10^{-4}
χ	Appendix C	2	10^2	2	69	0.91	0.0	4.72	0.81	4.2×10^{-4}

with a cumulative distribution

$$F(\delta x) = \frac{1 - e^{-N\delta x/\Delta X}}{1 - e^{-N}}, \qquad (B28)$$

normalizing over the range $0 \leq \delta x \leq \Delta x$. With N events, we expect the minimum spacing δx_{min} in our sample to probe the $1/N$ quantile of this exponential distribution. Setting $F = 1/N$ and inverting Eq. (B28) then gives

$$\delta x_{min} = -\frac{\Delta x}{N}\ln\left[1 - \frac{1}{N}(1 - e^{-N})\right]; \qquad (B29)$$

this is the minimum length scale we expect to probe with N events. We choose $\Sigma_{\ln\tau}$ such that 95% of our prior lies above $\ln\delta x_{min}$:

$$\Sigma_{\ln\tau} = \frac{\ln\delta x_{min} - \ln(\Delta x/2)}{\sqrt{2}\,\mathrm{Erf}^{-1}(1 - 2 \times 0.95)}. \qquad (B30)$$

So far, all of our discussion has concerned priors on the variance and length scale of an autoregressive process. We also need to consider a prior on the *mean* of the process, denoted above as $\ln r$. For all analyses, we place a logarithmically uniform prior on r,

$$p(r) = \frac{1}{r}, \qquad (B31)$$

across the interval $10^{-6} \leq (r/\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}) \leq 10^5$, with $p(r) = 0$ outside this interval.

Together, the product of Eqs. (B20), (B24), (B26), and (B31) comprise the priors placed on our autoregressive models, with variances defined by Eqs. (B21), (B25), and (B30), respectively. These priors still depend on several yet-undefined choices for Δx, $\delta\Psi_{max}$, and $\delta\Psi_{event}$, which

are listed in Table I along with several other derived quantities characterizing the above priors.

In Sec. II, we discussed how the likelihood is approximated via a weighted Monte Carlo average over ensembles of posterior samples; see Eq. (9). Similarly, the expected number of detections was evaluated by a Monte Carlo average over a set of successfully found injections; see Eq. (12). Both approximations break down if these averages become dominated by a very small number of posterior samples or found injections. One metric for gauging the health of Monte Carlo averaging is the effective sample number. Define $w_i(\Lambda) = R_d(\lambda_{inj,i}; \Lambda)/p_{pe}(\lambda_{inj,i})$ to be the weights appearing in the calculation of $N_{exp}(\Lambda)$ in Eq. (12). The effective number of samples informing this calculation is given by

$$N_{eff}^{inj}(\Lambda) = \frac{(\sum_i w_i(\Lambda))^2}{\sum_j (w_j(\Lambda))^2}. \qquad (B32)$$

In order for the systematic uncertainty in $N_{exp} = (\Lambda)$ to remain a subdominant effect in our hierarchical analysis, it is necessary that $N_{eff}^{inj}(\Lambda) \gtrsim 4N_{obs}$ [36,100]. Similarly, define $w_{I,j}(\Lambda) = R_d(\lambda_{I,j}; \Lambda)/p_{pe}(\lambda_{I,j})$ to be the weights defined in Eq. (9) over the posterior samples j of each event I. The number of effective posterior samples informing each event's likelihood is then

$$N_{eff,I}^{samp}(\Lambda) = \frac{(\sum_j w_{I,j}(\Lambda))^2}{\sum_k (w_{I,k}(\Lambda))^2}. \qquad (B33)$$

In particular, a useful metric is the minimum number of effective posterior samples, $\min_I N_{eff,I}^{samp}(\Lambda)$, taken over events I. Healthy inference generally requires $\min_I N_{eff,I}^{samp}(\Lambda) \gg 1$ [101].

For each of our analyses, we monitor $N_{\rm eff}^{\rm inj}$ and min $N_{\rm eff}^{\rm samp}$ for signs of poor effective sample counts; see Figs. 17–20 in Appendix E. Although potential population models with low effective sample counts are largely discouraged by the above priors on σ, $\ln \tau$, and $\sigma/\sqrt{\tau}$, we further prevent our inference from exploring models with pathologically low effective samples by severely penalizing proposed populations with $N_{\rm eff}^{\rm inj} < 4N_{\rm obs}$ and/or min $\log_{10} N_{\rm eff}^{\rm samp} < 0.6$. Specifically, we define the function

$$\mathcal{S}(x) = \left(\frac{1}{1 + x^{-30}}\right); \tag{B34}$$

this asymptotes to unity when x is large and falls to zero when x approaches zero. We then add the terms $\ln \mathcal{S}(N_{\rm eff}^{\rm inj}/4N_{\rm obs}) + \ln \mathcal{S}(\min \log_{10} N_{\rm eff}^{\rm samp}/0.6)$ to the log-likelihood implemented in numpyro. These terms send the log-likelihood towards $-\infty$ when either of the above conditions is violated.

APPENDIX C: DEMONSTRATIONS ON KNOWN POPULATIONS

We demonstrate the machinery developed in Appendices A and B by injecting and recovering a set of known distributions. This is useful in verifying that our methodology works as expected and in diagnosing any limitations in the performance of our autoregressive model or interpretation of our results. We consider four toy populations:

(1) A Gaussian distribution,

$$p(\chi) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\chi-\mu)^2/2\sigma^2}, \tag{C1}$$

with $\mu = 0.05$ and $\sigma = 1$.

(2) A delta function at zero,

$$p(\chi) = \delta(\chi). \tag{C2}$$

(3) A mixture between a delta function and a broad Gaussian,

$$p(\chi) = \frac{f_g}{\sqrt{2\pi\sigma^2}} e^{-(\chi-\mu)^2/2\sigma^2} + (1 - f_g)\delta(\chi), \tag{C3}$$

with $\mu = 0.5$, $\sigma = 0.3$, and $f_p = 0.3$.

(4) A half-normal distribution truncated to purely positive values,

$$p(\chi) = \begin{cases} \sqrt{\frac{2}{\pi\sigma^2}} e^{-(\chi-\mu)^2/2\sigma^2} & (\chi \geq 0) \\ 0 & (\chi < 0), \end{cases} \tag{C4}$$

with $\mu = 0$ and $\sigma = 0.35$.

All four are chosen as plausible distributions of the effective inspiral spin $\chi_{\rm eff}$ among binary black hole mergers. Specifically, the first model corresponds approximately to the measured effective spin distribution reported in Ref. [6]. The second and third correspond to cases in which some or all black hole spins are zero, as predicted in Ref. [66] and initially claimed in Ref. [57] (but disfavored by Refs. [15,16,60]). The fourth population corresponds to a situation in which there is no excess of zero-spin black holes but in which all spins are preferentially aligned and hence give purely positive $\chi_{\rm eff}$.

We draw 69 simulated events from each population, matching the size of the GWTC-3 catalog, and assume that these events are obtained over $T = 1$ yr of observation. For each simulated event i, we draw a "true" value $\chi_{{\rm true},i} \sim p(\chi)$ and then a random "observed" maximum likelihood value. To obtain an observed value, we first draw a random measurement uncertainty according to $\log_{10} \sigma_i \sim N(-0.9, 0.3)$, chosen to match the distribution of $\chi_{\rm eff}$ uncertainties among events in GWTC-3. We then randomly draw an observed value $\chi_{{\rm obs},i} \sim N(\chi_{{\rm true},i}, \sigma_i)$. Finally, we randomly draw 1000 mock posterior samples per event, assuming these are Gaussian distributed about $\chi_{{\rm obs},i}$ with standard deviation σ_i. For simplicity, we assume a flat selection function. To most accurately test the behavior of our method, however, we still compute the expected number of detections $N_{\rm exp}$ using a Monte Carlo average over a set of 10^4 discrete samples, as in Eq. (12). In particular, this means that our inference can still experience the instability discussed above in Appendix B. The final line in Table I defines the prior placed on our autoregressive model in each of the four cases.

Our results are shown in Fig. 16. For each injected population, the left-hand column shows our recovery of the normalized probability distribution, while the right-hand column shows inferred rate densities $dR/d\chi$. In all cases, the underlying distributions from which mock events are drawn are shown via dashed red curves. Case 1 shows faithful recovery of a Gaussian distribution, with the reconstructed probability distribution and rate density closely matching the injected population within the range $-0.5 \lesssim \chi \lesssim 0.5$. Beyond this range, we see that, although our recovered rate density is able to fall by approximately 3 orders from the center to the edge of the domain, it is unable to go strictly to zero. Instead, the autoregressive process asymptotes to $dR/d\chi \approx 0.5$ in regions with no data. Note that this asymptotic limit corresponds to the threshold below which we expect no additional events on top of our 69 simulated detections after $T = 1$ yr of observation. As our catalog increases in size, the continued nonobservation of events at $|\chi| \gtrsim 0.5$ will further suppress tails in the recovered rate density.

Case 2 behaves similarly. Although our autoregressive model cannot truly produce a delta function distribution, we recover a narrow distribution centered at $\chi = 0$.
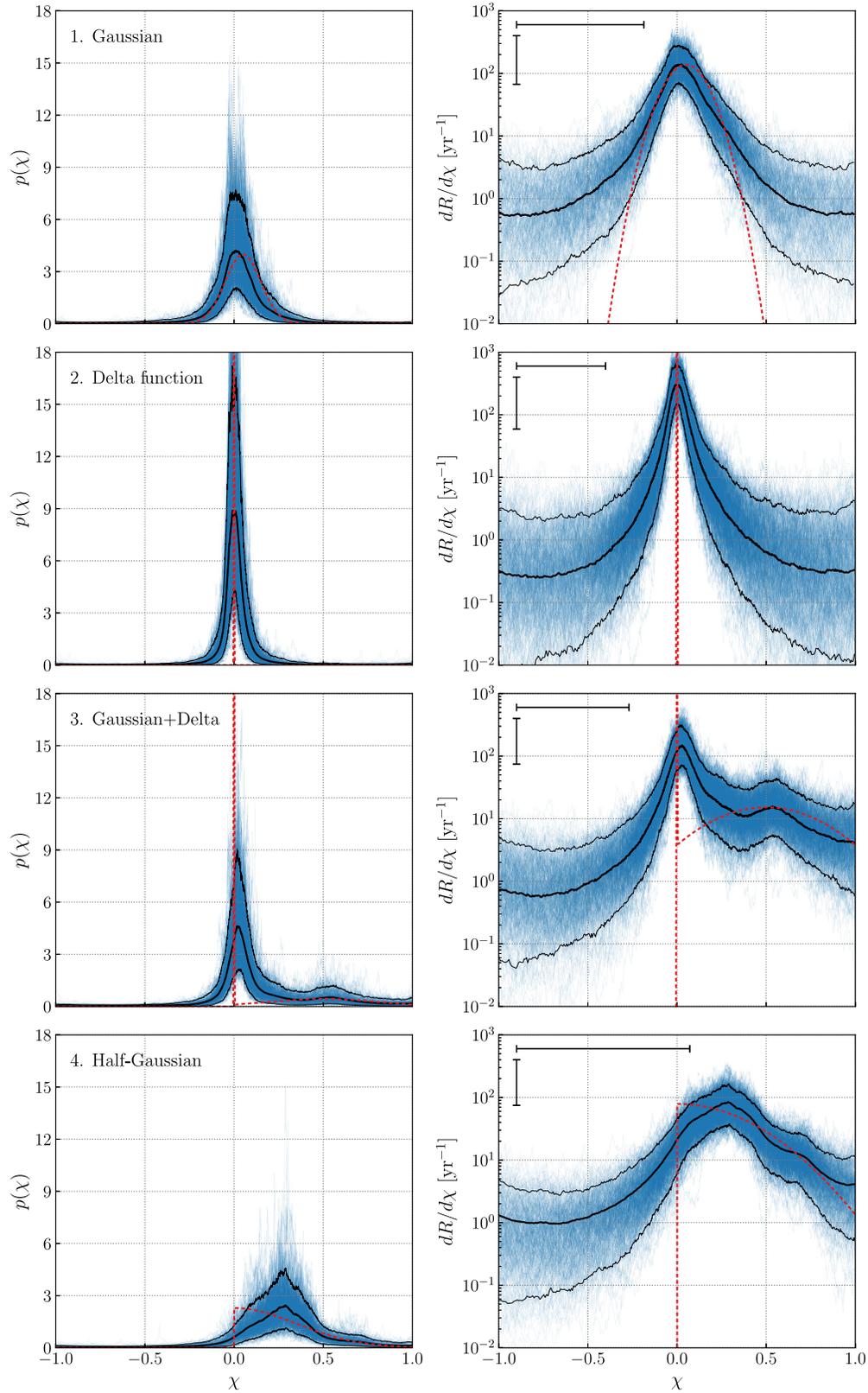
FIG. 16. Demonstrated autoregressive inference of four simulated populations, as described in Appendix C. The left column indicates the normalized probability distributions recovered in each case, while the right column shows inferred rate densities. In the right-hand column, the median recovered autocorrelation length τ and standard deviation e^σ of the process are shown via the horizontal and vertical error bars (recall that σ^2 defines the variance of the *logarithmic* number density). In each case, true distributions are indicated as dashed red curves.

Once more, we see that the inferred rate falls by 3 orders of magnitude between $\chi = 0$ and $|\chi| \gtrsim 0.5$, with asymptotic tails in regions with no data. In Case 4, illustrating recovery of a broad half-Gaussian distribution, we correctly recover the injected rate density from its peak of $dR/d\chi \approx 10^2$ at $\chi = 0$ down to $dR/d\chi \approx 1$ at $\chi = 0$. At negative χ, Case 4 shows the same asymptotic tail exhibited by Cases 1 and 2.

One limitation of our autoregressive model is illustrated in Case 3, in which our injected population is a mixture between a delta function at $\chi = 0$ and a broad Gaussian at positive χ. We see that both features, a primary peak at $\chi = 0$ and a secondary peak at $\chi = 0.5$, are correctly and confidently identified by our inference despite the order-of-magnitude difference in the prominence of these features. The rate density of events arising from the half-Gaussian, in particular, is well recovered even out to $\chi = 1$. At the same time, our autoregressive model struggles to produce an extremely sharp feature at $\chi = 0$ while *also* fitting the broad Gaussian. Recovery of the Gaussian serves to fix the autocorrelation length τ of the process to reasonably large values, in turn limiting the resolution with which the delta function can be resolved. Compare, for example, the recovery of the delta function in Case 3 to the much sharper recovery of an isolated delta function in Case 2.

APPENDIX D: STRONGLY PARAMETRIZED MODELS

In the main text, we studied the distributions of black hole masses (Sec. III), spins (Secs. V and VI), and redshifts (Sec. IV), modeling each set of distributions, in turn, as autoregressive processes. To accurately measure the population distribution of any one parameter, it is generally necessary to simultaneously fit for the distributions of other parameters. Therefore, wherever we focused on modeling a specific subset of parameters using autoregressive models, we concurrently fit the remaining parameters using simple strongly parametrized models. The priors used for each of the following models are given in Table II.

Our parametric mass model assumes that primary masses are drawn from a mixture between a power law and a Gaussian peak, with possible tapering at low and high

TABLE II. Parameters characterizing the ordinary strongly parametrized models discussed in Appendix D. For each parameter, we indicate its defining equation, the sections in which it is used, and its associated prior. We use $N(a, b)$ to denote a Gaussian prior with mean a and standard deviation b, $U(a, b)$ to indicate a uniform prior between a and b, and $LU(a, b)$ to indicate a logarithmically uniform prior between the given bounds.

Parameter	Defined	Sections	Prior
λ	Eq. (D1)	IV,V,VI	$N(0, 10)$
μ_m	Eq. (D1)	IV,V,VI	$U(20M_\odot, 50M_\odot)$
σ_m	Eq. (D1)	IV,V,VI	$U(2M_\odot, 15M_\odot)$
f_p	Eq. (D1)	IV,V,VI	$LU(10^{-3}, 1)$
$m_{\rm low}$	Eq. (D2)	IV,V,VI	$U(5M_\odot, 15M_\odot)$
$m_{\rm high}$	Eq. (D2)	IV,V,VI	$U(50M_\odot, 100M_\odot)$
$\delta m_{\rm low}$	Eq. (D2)	IV,V,VI	$LU(0.1M_\odot, 10M_\odot)$
$\delta m_{\rm high}$	Eq. (D2)	IV,V,VI	$LU(3M_\odot, 30M_\odot)$
β_q	Eq. (D3)	IV,V,VI	$N(0, 4)$
μ_χ	Eq. (D4)	III,IV	$U(0, 1)$
σ_χ	Eq. (D4)	III,IV	$LU(0.1, 1)$
σ_u	Eq. (D5)	III,IV	$U(0.3, 2)$
κ	Eq. (D6)	III,V,VI	$N(0, 5)$

masses. This model is a variant of the POWER LAW+PEAK model first defined in Ref. [10] and used in depth in Refs. [6,9]. Specifically, we define

$$\phi(m_1) = \frac{f_p}{\sqrt{2\pi\sigma_m^2}}\exp\left(-\frac{(m_1 - \mu_m)^2}{2\sigma_m^2}\right)$$
$$+ (1 - f_p)\left(\frac{1 + \lambda}{(100M_\odot)^{1+\lambda} - (2M_\odot)^{1+\lambda}}\right)m_1^\lambda \quad \text{(D1)}$$

to be the superposition of a power law and Gaussian; the former is normalized in the range $2M_\odot \leq m_1 \leq 100M_\odot$ with spectral index λ, while the latter is centered at mean μ_m with standard deviation σ_m. The parameter f_p controls the relative contribution of each component. Our complete primary mass distribution is of the shape

$$f(m_1) = \begin{cases} \phi(m_1)\exp\left[\frac{-(m_1 - m_{\rm low})^2}{2\delta m_{\rm low}^2}\right] & (m_1 < m_{\rm low}) \\ \phi(m_1) & (m_{\rm low} \leq m_1 \leq m_{\rm high}) \\ \phi(m_1)\exp\left[\frac{-(m_1 - m_{\rm high})^2}{2\delta m_{\rm high}^2}\right] & (m_{\rm high} < m_1), \end{cases} \quad \text{(D2)}$$

with squared exponentials that taper $f(m_1)$ towards zero above and below $m_{\rm low}$ and $m_{\rm high}$, respectively. The tapering scales $\delta m_{\rm low}$ and $\delta m_{\rm high}$ are additional free parameters inferred from the data. Mass ratios, in turn, are assumed to follow a power-law distribution, with

$$p(q|m_1) = \left(\frac{1 + \beta_q}{m_1^{1+\beta_q} - (2M_\odot)^{1+\beta_q}}\right)m_2^{\beta_q}. \quad \text{(D3)}$$

This parametric mass model is used in Secs. IV–VI, and when focusing on autoregressive modeling of black hole spins and redshifts.

When nonparametrically exploring the black hole mass and redshift distributions, we revert to a parametric spin model in which component spin magnitudes and spin-orbit tilt angles are independently and identically distributed as truncated normal distributions. Each component spin magnitude in a given binary has a probability distribution

$$p(\chi_i) = \sqrt{\frac{2}{\pi\sigma_\chi^2}} \frac{e^{-(\chi_i-\mu_\chi)^2/2\sigma_\chi^2}}{\text{Erf}\left(\frac{1-\mu_\chi}{\sqrt{2\sigma_\chi^2}}\right) + \text{Erf}\left(\frac{\mu_\chi}{\sqrt{2\sigma_\chi^2}}\right)}, \quad (D4)$$

with a mean μ_χ and standard deviation σ_χ that are inferred from the data. The cosines of component spin-tilt angles, meanwhile, are independently distributed as

$$p(\cos\theta_i) = \sqrt{\frac{2}{\pi\sigma_u^2}} \frac{e^{-(\cos\theta_i-1)^2/2\sigma_u^2}}{\text{Erf}\left(\frac{-2}{\sqrt{2\sigma_u^2}}\right)}, \quad (D5)$$

with a mean fixed to 1 but a standard deviation σ_u measured from the data. This parameteric model is used in Secs. III and IV when targeting binary masses and redshifts with our autoregressive prior.

Finally, when targeting binary masses or spins, we revert to a standard parametric redshift model in which the comoving merger rate density grows as

$$\mathcal{R}(\theta; z) \propto (1+z)^\kappa. \quad (D6)$$

The observed detector-frame merger rate per unit redshift correspondingly grows as

$$R(z) \propto (1+z)^{\kappa-1}\left(\frac{dV_c}{dz}\right), \quad (D7)$$

where the additional factor of $(1+z)^{-1}$ converts between source- and detector-frame rates. This parametric model is adopted in Secs. III, V, and VI when nonparametrically measuring the black hole mass and spin distributions.

APPENDIX E: ADDITIONAL POSTERIORS AND PREDICTIVE CHECKS

In this appendix, we show a few more supplemental results that may be useful in assessing the behavior and performance of our population inference with an autoregressive model.

As discussed elsewhere, our autoregressive process models are characterized by hyperparameters σ and τ controlling the variance and autocorrelation length of the (log) merger rate. Figures 17–20 show the posteriors obtained on these parameters for each autoregressive model used in the main text. For diagnostic purposes, these figures also show quantities related to the effective number of

samples informing our inference. Included in each corner plot is the distribution of effective injections per observation, $N_{\text{eff}}^{\text{inj}}/N_{\text{obs}}$, and the distribution of effective posterior sample counts, minimized across events. See Appendix B for further details regarding these diagnostics.

Furthermore, we assess the validity of our results using posterior predictive checks, comparing distributions of observed binary parameters to distributions predicted by our fitted models. Catalogs of "observed" and "predicted" parameters are generated by resampling individual event posteriors as well as the set of injections used elsewhere to compute $N_{\text{exp}}(\Lambda)$. This process proceeds as follows [102]:

(1) Randomly draw a sample Λ_i from our posterior on the population parameters Λ (this includes the variables characterizing strongly parametrized models, as well as latent parameters defining any autoregressive processes in use).

(2) For each observed event, with posterior samples $\{\lambda_j\}$, compute weights $w_j = R_d(\lambda_j|\Lambda_i)/p_{\text{pe}}(\lambda_j)$, the ratios between the detector-frame distribution defined by Λ_i and the parameter estimation prior $p_{\text{pe}}(\lambda)$.

(3) From each observed event, draw a single sample λ_j with draw probabilities proportional to the weights $\{w_j\}$. The resulting set $\{\lambda\}$ defines a single catalog of "observed" parameters consistent with our data.

(4) For each successfully recovered injection, with parameters $\{\lambda_{\text{inj},j}\}$, compute weights $w_j = R_d(\lambda_{\text{inj},j}|\Lambda_i)/p_{\text{inj}}(\lambda_{\text{inj},j})$, where $p_{\text{inj}}(\lambda)$ is the distribution from which injections were drawn.

(5) Draw N_{obs} values $\lambda_{\text{inj},j}$ with draw probabilities proportional to the above weights. The resulting set $\{\lambda_{\text{inj}}\}$ defines a single catalog of "predicted" parameters consistent with the proposed population Λ_i and with appropriate selection effects.

(6) Sort the "observed" and "predicted" catalogs, and plot against one another.

(7) Repeat.

Figure 21 shows the result of this algorithm using the population model adopted in Sec. III, in which the primary mass and mass ratio distributions of binary black holes are described as autoregressive processes. For convenience, the dotted black lines mark the diagonals along which "observed" and "predicted" values are equal. A systematic deviation from this diagonal would indicate tension between our model (and its corresponding predictions) and our observed data. In Fig. 21, all traces are systematically clustered around the diagonal, with no indication of systematic mismodeling. Figures 22–24 similarly show predictive checks on the autoregressive models used in Secs. IV–VI to describe distributions of binary redshifts, component spins, and effective spins; they also indicate good agreement between our autoregressive models and observed data.
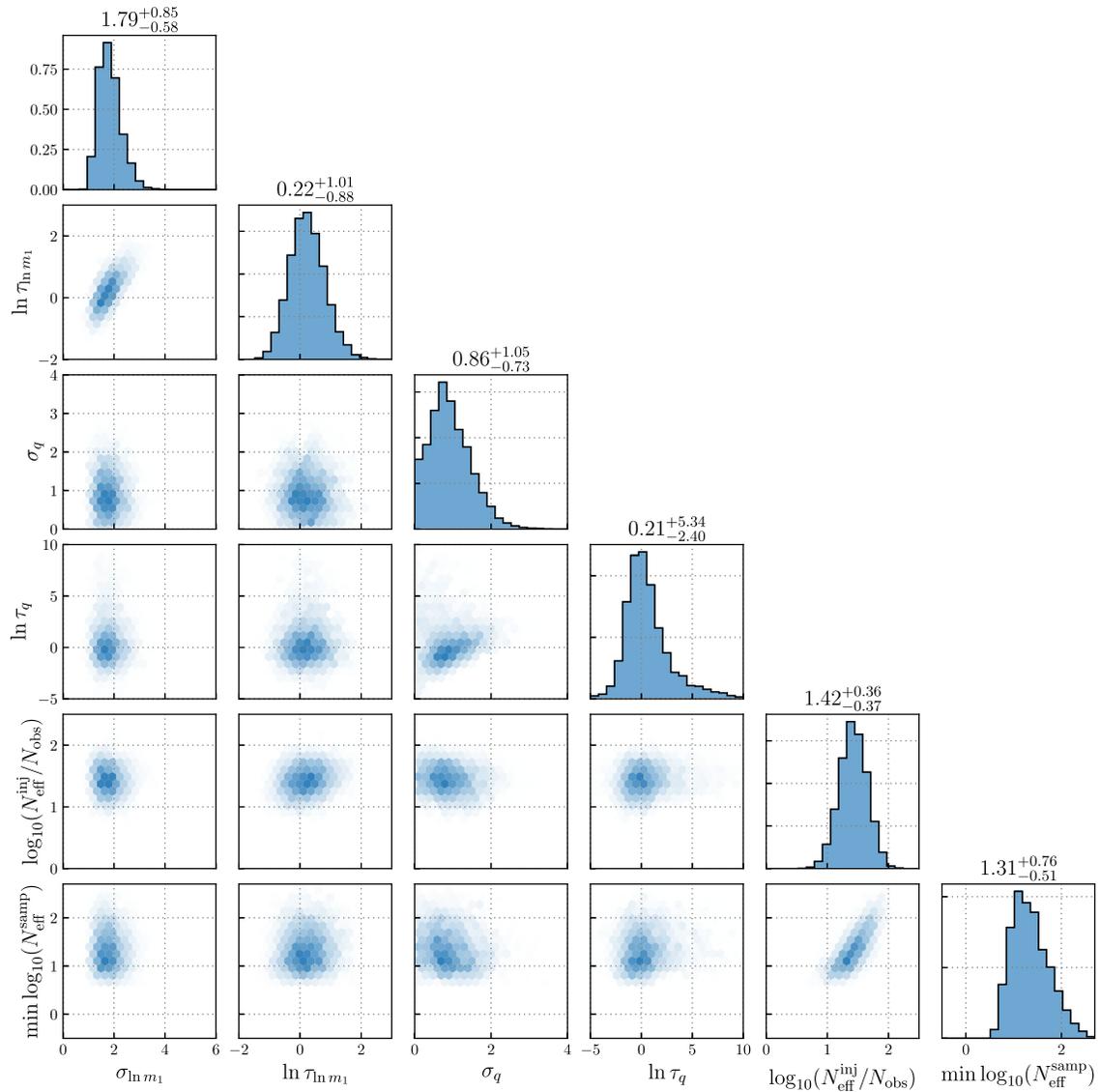
FIG. 17. Posteriors on the hyperparameters characterizing the primary mass and mass-ratio distributions presented in Sec. III. As a diagnostic, we also include the effective number of injections per observed event and the effective number of posterior samples informing the per-event likelihood, minimized over the 69 events in our sample.

In addition to checking model validity, the catalogs of "predicted" detections enable other related calculations regarding the locations in which detections are expected (or not expected) to arise. In particular, we saw in Appendix C that autoregressive processes can exhibit prior-dominated tails in regions with no data and that the onset of these tails can be characterized by identifying regions in which fewer than one detection is predicted by

the fitted model. Accordingly, we can use the ensembles of "predicted" detections appearing in Figs. 21–24 to assess the statistical significance of tails appearing in our autoregressive results, calculating threshold values above or below which fewer than N detections are predicted by our fitted models, on average. The results of this calculation are quoted in Secs. V and VI in the main text, when discussing the robustness of our recovered spin distributions.
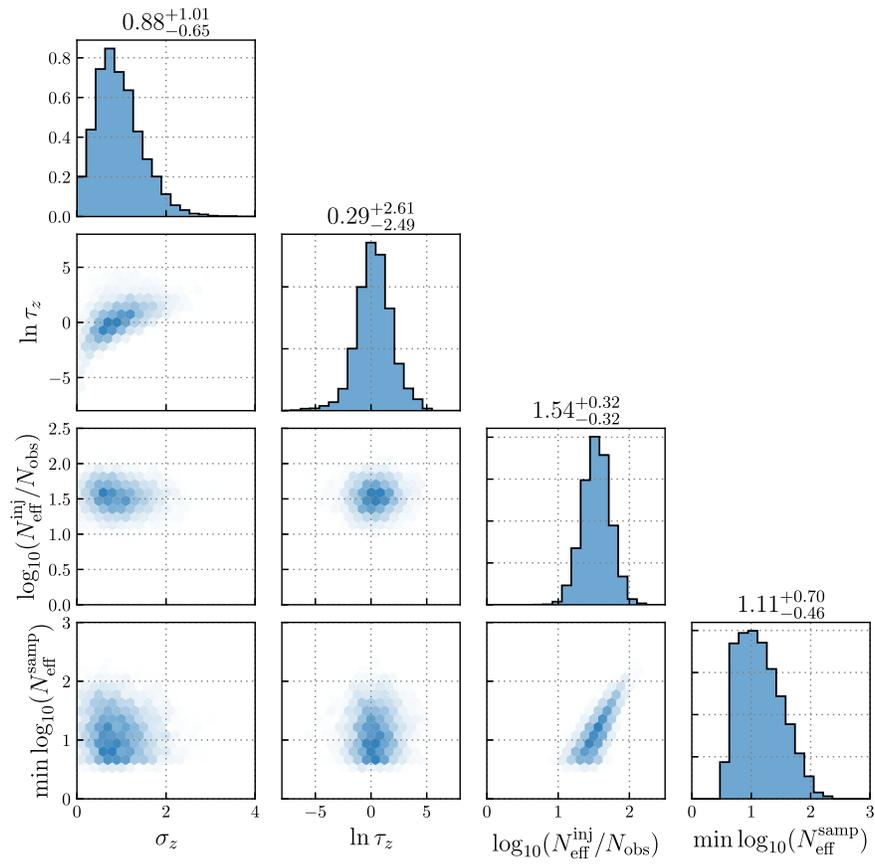
FIG. 18.   As in Fig. 17 but for the redshift evolution of the merger rate shown in Sec. IV.
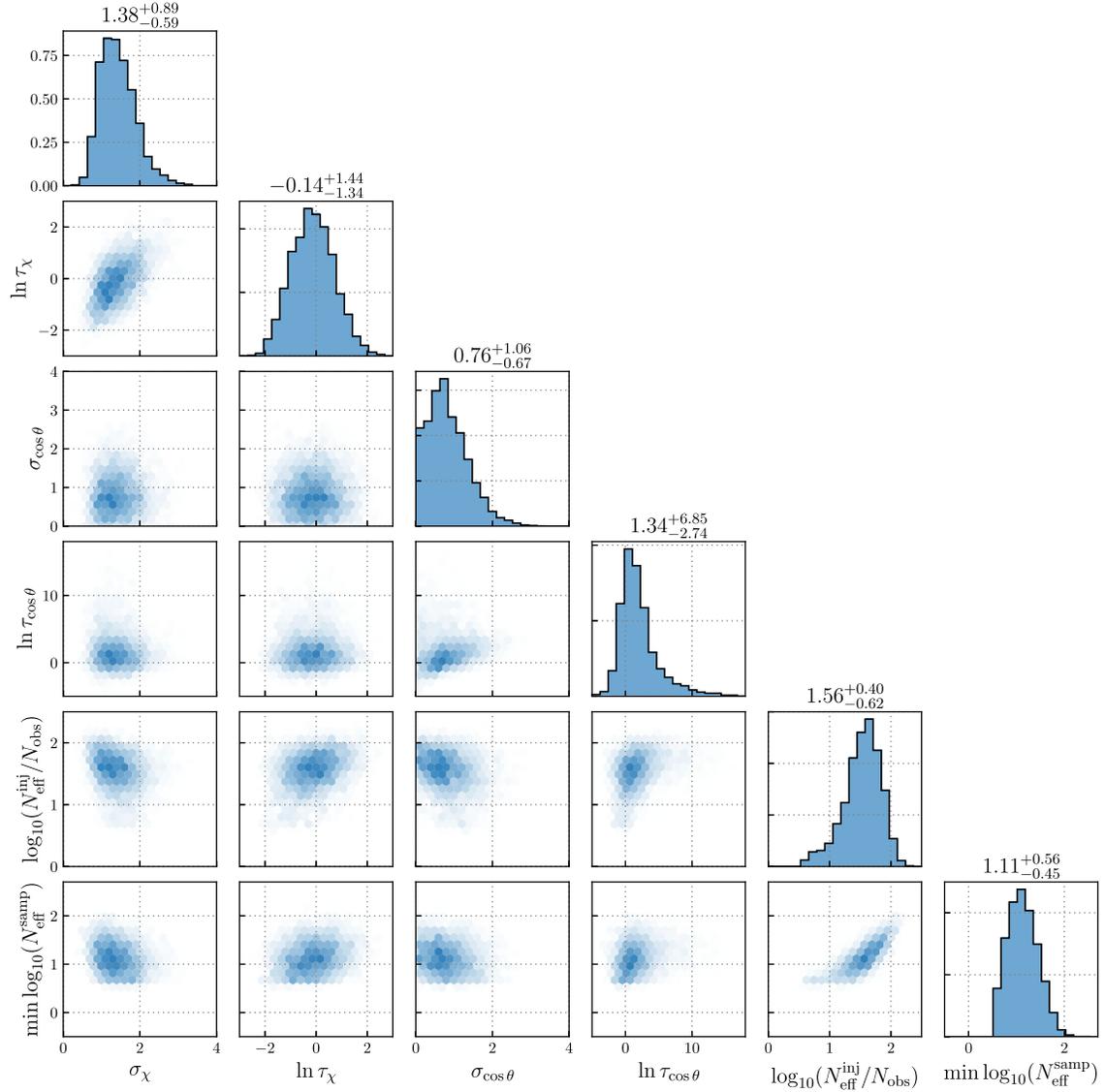
FIG. 19.   As in Fig. 17 but for the component spin distributions shown in Sec. V.
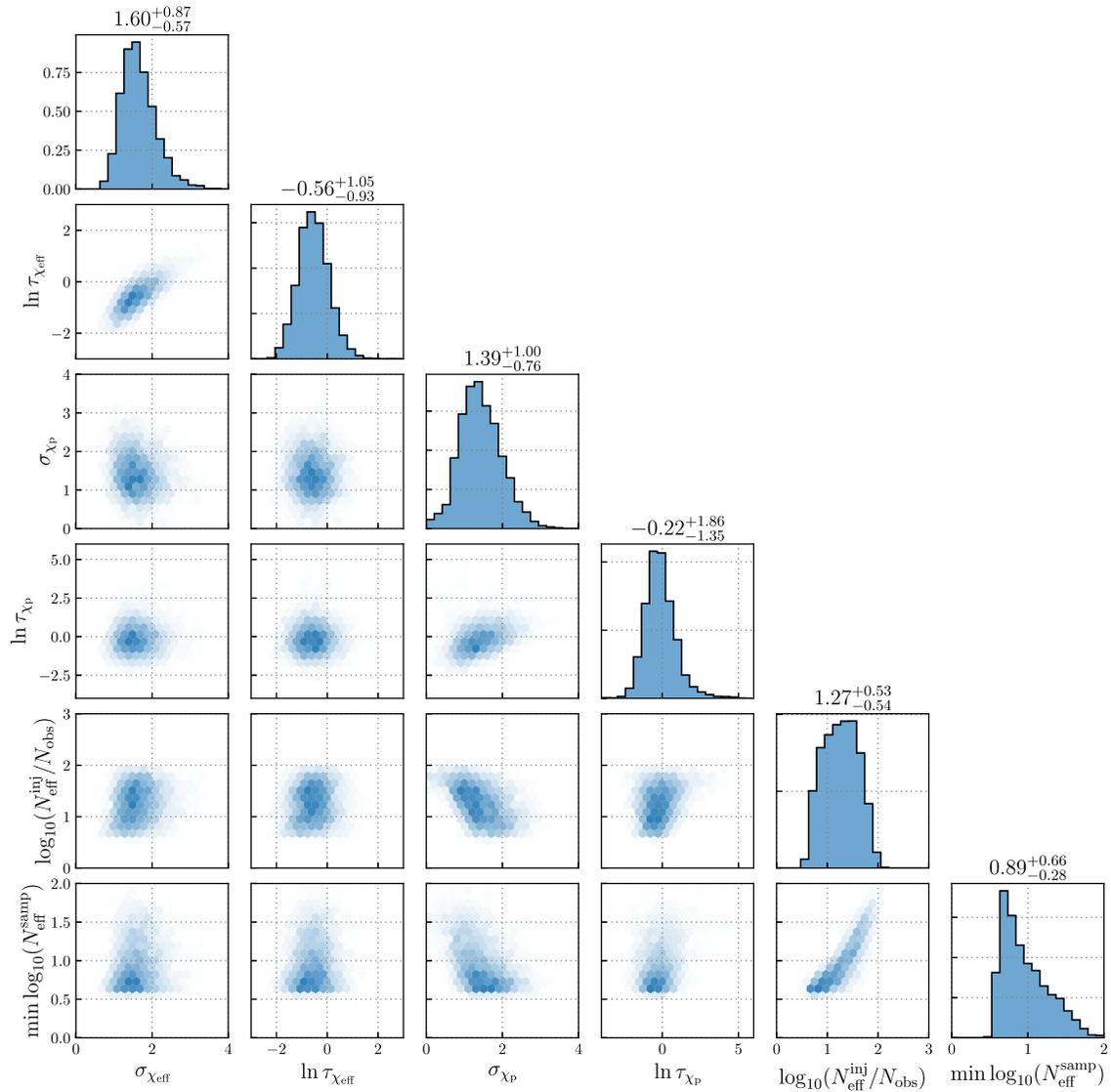
FIG. 20.   As in Fig. 17 but for the effective spin distributions shown in Sec. VI.
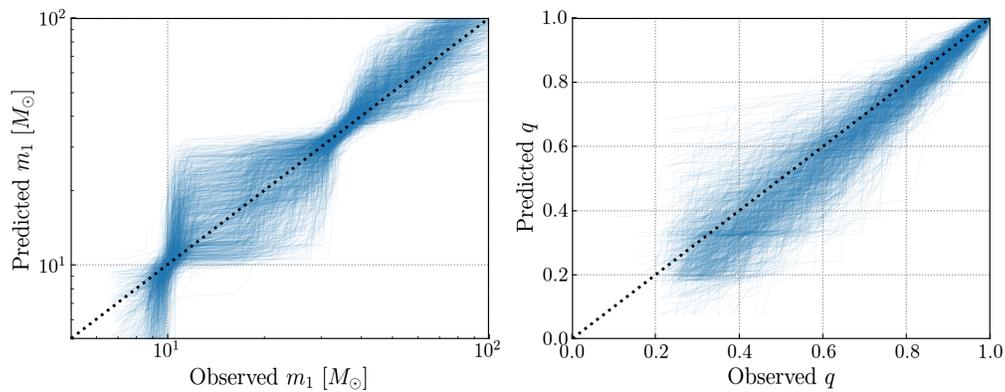

FIG. 21.   Posterior predictive check on primary masses and mass ratios, according to the autoregressive model over $\ln m_1$ and q presented in Sec. III. Each trace corresponds to an individual pass through the algorithm described in Appendix E. In both cases, traces are centered along the diagonal (marked with a dotted black line), indicating no tension between the fitted population model and our observed data.
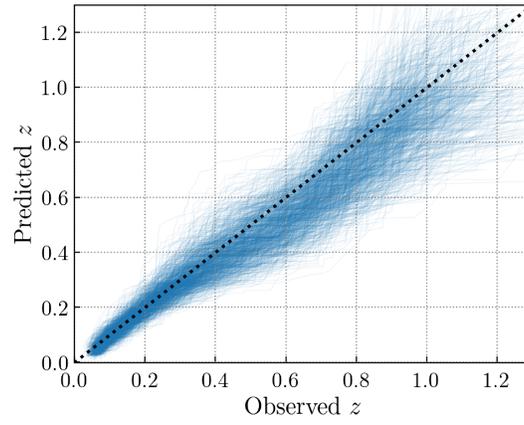
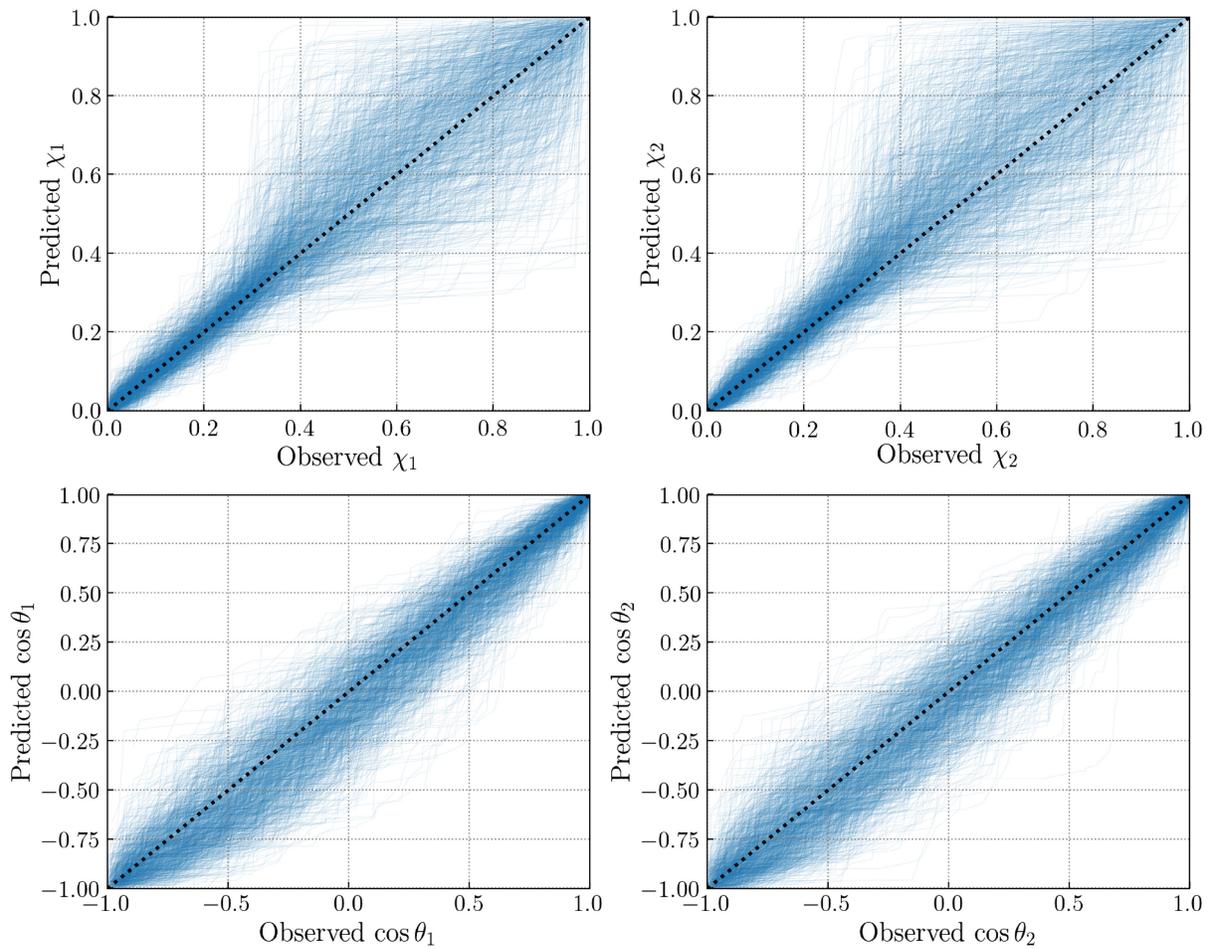FIG. 22. As in Fig. 21 but for the redshift distribution modeled as an autoregressive process in Sec. IV.

FIG. 23. As in Fig. 21 but for the component spin magnitude and tilt distributions modeled as autoregressive process in Sec. V. We show predictive checks on both component spins to gauge any inconsistency with our assumptions regarding component spin independence; see Eq. (29).
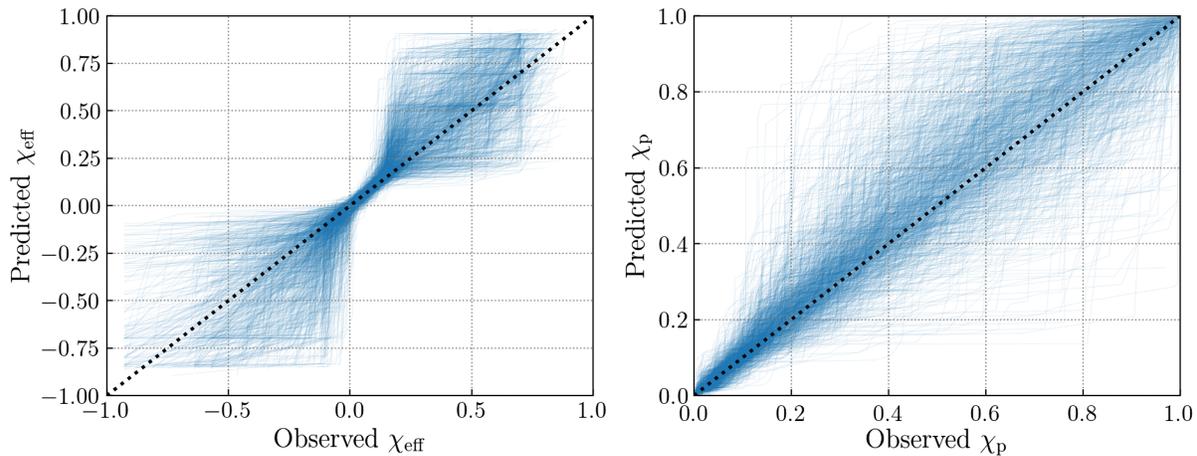
FIG. 24. As in Fig. 21 but for the effective spin distributions modeled as an autoregressive process in Sec. VI.

[1] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run*, Phys. Rev. X **13**, 041039 (2023).

[2] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy *et al.* (LIGO Scientific Collaboration), *Advanced LIGO*, Classical Quantum Gravity **32**, 074001 (2015).

[3] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou *et al.* (Virgo Collaboration), *Advanced Virgo: A second-generation interferometric gravitational wave detector*, Classical Quantum Gravity **32**, 024001 (2015).

[4] T. Akutsu, M. Ando, K. Arai, Y. Arai, S. Araki *et al.* (KAGRA Collaboration), *Overview of KAGRA: Detector design and construction history*, Prog. Theor. Exp. Phys. (2021) 05A101.

[5] Counting those events with false alarm rates FAR < 1 yr^{-1} [6].

[6] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *Population of merging compact binaries inferred using gravitational waves through GWTC-3*, Phys. Rev. X **13**, 011048 (2023).

[7] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, *4-OGC: Catalog of gravitational waves from compact-binary mergers*, Astrophys. J. **946**, 59 (2023).

[8] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, *New binary black hole mergers in the LIGO-Virgo O3a data*, Phys. Rev. D **106**, 043009 (2022).

[9] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Population properties of compact objects from the second LIGO–Virgo Gravitational-Wave Transient Catalog*, Astrophys. J. Lett. **913**, L7 (2021).

[10] C. Talbot and E. Thrane, *Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization*, Astrophys. J. **856**, 173 (2018).

[11] D. Wysocki, J. Lange, and R. O'Shaughnessy, *Reconstructing phenomenological distributions of compact binaries via gravitational wave observations*, Phys. Rev. D **100**, 043012 (2019).

[12] B. Edelman, Z. Doctor, and B. Farr, *Poking holes: Looking for gaps in LIGO/Virgo's black hole population*, Astrophys. J. Lett. **913**, L23 (2021).

[13] V. Tiwari, *Exploring features in the binary black hole population*, Astrophys. J. **928**, 155 (2022).

[14] A. Farah, M. Fishbach, R. Essick, D. E. Holz, and S. Galaudage, *Bridging the gap: Categorizing gravitational-wave events at the transition between neutron stars and black holes*, Astrophys. J. **931**, 108 (2022).

[15] T. A. Callister, S. J. Miller, K. Chatziioannou, and W. M. Farr, *No evidence that the majority of black holes in binaries have zero spin*, Astrophys. J. Lett. **937**, L13 (2022).

[16] M. Mould, D. Gerosa, F. S. Broekgaarden, and N. Steinle, *Which black hole formed first? Mass-ratio reversal in massive binary stars from gravitational-wave data*, Mon. Not. R. Astron. Soc. **517**, 2738 (2022).

[17] T. A. Callister, C.-J. Haster, K. K. Y. Ng, S. Vitale, and W. M. Farr, *Who ordered that? Unequal-mass binary black hole mergers have larger effective spins*, Astrophys. J. Lett. **922**, L5 (2021).

[18] C. Adamcewicz and E. Thrane, *Do unequal-mass binary black hole systems have larger χ_{eff}? Probing correlations with copulas in gravitational-wave astronomy*, Mon. Not. R. Astron. Soc. **517**, 3928 (2022).

[19] B. Edelman, Z. Doctor, J. Godfrey, and B. Farr, *Ain't no mountain high enough: Semiparametric modeling of LIGO-Virgo's binary black hole mass distribution*, Astrophys. J. **924**, 101 (2022).

[20] B. Edelman, B. Farr, and Z. Doctor, *Cover your basis: Comprehensive data-driven characterization of the binary black hole population*, Astrophys. J. **946**, 16 (2023).

[21] J. Golomb and C. Talbot, *Searching for structure in the binary black hole spin distribution*, Phys. Rev. D **108**, 103009 (2023).

[22] I. Mandel, W. M. Farr, A. Colonna, S. Stevenson, P. Tiňo, and J. Veitch, *Model-independent inference on compact-binary observations*, Mon. Not. R. Astron. Soc. **465**, 3254 (2017).

[23] B. Farr, D. E. Holz, and W. M. Farr, *Using spin to understand the formation of LIGO and Virgo's black holes*, Astrophys. J. **854**, L9 (2018).

[24] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Binary black hole population properties inferred from the first and second observing runs of Advanced LIGO and Advanced Virgo*, Astrophys. J. **882**, L24 (2019).

[25] M. Fishbach, W. M. Farr, and D. E. Holz, *The most massive binary black hole detections and the identification of population outliers*, Astrophys. J. **891**, L31 (2020).

[26] D. Veske, I. Bartos, Z. Márka, and S. Márka, *Characterizing the observation bias in gravitational-wave detections and finding structured population properties*, Astrophys. J. **922**, 258 (2021).

[27] V. Tiwari, *VAMANA: Modeling binary black hole population with minimal assumptions*, Classical Quantum Gravity **38**, 155007 (2021).

[28] S. Rinaldi and W. Del Pozzo, *(H)DPGMM: A hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function*, Mon. Not. R. Astron. Soc. **509**, 5454 (2022).

[29] J. Powell, S. Stevenson, I. Mandel, and P. TiÅo, *Unmodelled clustering methods for gravitational wave populations of compact binary mergers*, Mon. Not. R. Astron. Soc. **488**, 3810 (2019).

[30] J. Sadiq, T. Dent, and D. Wysocki, *Flexible and fast estimation of binary merger population distributions with an adaptive kernel density estimator*, Phys. Rev. D **105**, 123014 (2022).

[31] T. J. Loredo, *Accounting for source uncertainties in analyses of astronomical survey data*, AIP Conf. Proc. **735**, 195 (2004).

[32] S. R. Taylor and D. Gerosa, *Mining gravitational-wave catalogs to understand binary stellar evolution: A new hierarchical Bayesian framework*, Phys. Rev. D **98**, 083017 (2018).

[33] I. Mandel, W. M. Farr, and J. R. Gair, *Extracting distribution parameters from multiple uncertain observations with selection biases*, Mon. Not. R. Astron. Soc. **486**, 1086 (2019).

[34] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, *Inferring the properties of a population of compact binaries in presence of selection effects*, in *Handbook of Gravitational Wave Astronomy* (Springer, Singapore, 2020).

[35] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, *GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run—O1 + O2 + O3 search sensitivity estimates*, Zenodo (2021), 10.5281/zenodo .5636816.

[36] W. M. Farr, *Accuracy requirements for empirically measured selection functions*, Res. Notes Am. Astron. Soc. **3**, 66 (2019).

[37] A. M. Farah, B. Edelman, M. Zevin, M. Fishbach, J. María Ezquiaga, B. Farr, and D. E. Holz, *Things that might go bump in the night: Assessing structure in the binary black hole mass spectrum*, Astrophys. J. **955**, 107 (2023).

[38] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary *et al.*, *JAX: Composable transformations of Python + NumPy programs* (2018).

[39] D. Phan, N. Pradhan, and M. Jankowiak, *Composable effects for flexible and accelerated probabilistic programming in numpyro*, arXiv:1912.11554.

[40] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan *et al.*, *Pyro: Deep universal probabilistic programming*, J. Mach. Learn. Res. **20**, 28 (2019), https://jmlr.org/papers/v20/18-403.html.

[41] M. D. Hoffman and A. Gelman, *The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo*, J. Mach. Learn. Res. **15**, 1593 (2014), https://jmlr.org/papers/v15/hoffman14a.html.

[42] M. Betancourt, *A conceptual introduction to Hamiltonian Monte Carlo*, arXiv:1701.02434.

[43] M. Fishbach, D. E. Holz, and W. M. Farr, *Does the black hole merger rate evolve with redshift?*, Astrophys. J. **863**, L41 (2018).

[44] For convenience, we neglect the full functional dependence $\mathcal{R}(\ln m_1, q, \chi_1, \chi_2, \cos\theta_1, \cos\theta_2; z)$ and just abbreviate this quantity as $\mathcal{R}(\ln m_1)$ since we are only concerned with the dependence of the merger rate on mass at the moment. We will use an analogous shorthand below when focusing on other parameters as well.

[45] V. Tiwari and S. Fairhurst, *The emergence of structure in the binary black hole mass distribution*, Astrophys. J. Lett. **913**, L19 (2021).

[46] M. Fishbach and D. E. Holz, *Where are LIGO's big black holes?*, Astrophys. J. **851**, L25 (2017).

[47] E. J. Baxter, D. Croon, S. D. McDermott, and J. Sakstein, *Find the gap: Black hole population analysis with an astrophysically motivated mass function*, Astrophys. J. Lett. **916**, L16 (2021).

[48] M. Fishbach, Z. Doctor, T. Callister, B. Edelman, J. Ye, R. Essick, W. M. Farr, B. Farr, and D. E. Holz, *When are LIGO/Virgo's big black hole mergers?*, Astrophys. J. **912**, 98 (2021).

[49] S. E. Woosley and A. Heger, *The pair-instability mass gap for black holes*, Astrophys. J. Lett. **912**, L31 (2021).

[50] Within Ref. [6], the right-hand panel of Fig. 10 appears to be in tension with this significance estimate, instead showing a measurement of $\mathcal{R}(q)$ (marginalized over all m_1) that unambiguously increases as a function of q. The behavior in Fig. 10 is actually due to the presumed *truncation* in the mass ratio distribution, rather than a confident measurement of positive β_q. Since the overall merger rate is highest at small m_1, the structure of the marginalized $\mathcal{R}(q)$ must correspondingly be dominated by the mass ratio distribution at small m_1. The truncation in $p(q|m_1)$, however, enforces that $q \approx 1$ when m_1 is small. This combination of effects requires $\mathcal{R}(q)$ to be maximized

at $q = 1$ after marginalization over m_1, nearly independently of β_q.

[51] T. Callister, M. Fishbach, D. E. Holz, and W. M. Farr, *Shouts and murmurs: Combining individual gravitational-wave sources with the stochastic background to measure the history of binary black hole mergers*, Astrophys. J. **896**, L32 (2020).

[52] L. A. C. van Son, S. E. de Mink, T. Callister, S. Justham, M. Renzo, T. Wagg, F. S. Broekgaarden, F. Kummer, R. Pakmor, and I. Mandel, *The redshift evolution of the binary black hole merger rate: A weighty matter*, Astrophys. J. **931**, 17 (2022).

[53] E. Payne and E. Thrane, *Model exploration in gravitational-wave astronomy with the maximum population likelihood*, Phys. Rev. Res. **5**, 023013 (2023).

[54] S. Biscoveanu, T. A. Callister, C.-J. Haster, K. K. Y. Ng, S. Vitale, and W. M. Farr, *The binary black hole spin distribution likely broadens with redshift*, Astrophys. J. Lett. **932**, L19 (2022).

[55] V. Varma, S. Biscoveanu, M. Isi, W. M. Farr, and S. Vitale, *Hints of spin-orbit resonances in the binary black hole population*, Phys. Rev. Lett. **128**, 031101 (2022).

[56] J. Roulet, H. S. Chia, S. Olsen, L. Dai, T. Venumadhav, B. Zackay, and M. Zaldarriaga, *Distribution of effective spins and masses of binary black holes from the LIGO and Virgo O1-O3a observing runs*, Phys. Rev. D **104**, 083010 (2021).

[57] S. Galaudage, C. Talbot, T. Nagar, D. Jain, E. Thrane, and I. Mandel, *Building better spin models for merging binary black holes: Evidence for nonspinning and rapidly spinning nearly aligned subpopulations*, Astrophys. J. Lett. **921**, L15 (2021).

[58] A subsequent erratum [59] diminished the initial evidence in Ref. [57] for distinct nonspinning and spinning subpopulations, bringing their conclusions into closer agreement with those of Refs. [6,15,16,60].

[59] S. Galaudage, C. Talbot, T. Nagar, D. Jain, E. Thrane, and I. Mandel, *Building better spin models for merging binary black holes: Evidence for nonspinning and rapidly spinning nearly aligned subpopulations*, Astrophys. J. Lett. **921**, L15 (2021); **936**, L18(E) (2022).

[60] H. Tong, S. Galaudage, and E. Thrane, *Population properties of spinning black holes using the Gravitational-wave Transient Catalog 3*, Phys. Rev. D **106**, 103019 (2022).

[61] Y. Qin, T. Fragos, G. Meynet, J. Andrews, M. Sørensen, and H. F. Song, *The spin of the second-born black hole in coalescing binary black holes*, Astron. Astrophys. **616**, A28 (2018).

[62] S. S. Bavera, T. Fragos, Y. Qin, E. Zapartas, C. J. Neijssel, I. Mandel, A. Batta, S. M. Gaebel, C. Kimball, and S. Stevenson, *The origin of spin in binary black holes: Predicting the distributions of the main observables of Advanced LIGO*, Astron. Astrophys. **635**, A97 (2020).

[63] K. Belczynski, J. Klencki, C. E. Fields, A. Olejak, E. Berti *et al.*, *Evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates for LIGO/Virgo binary black holes*, Astron. Astrophys. **636**, A104 (2020).

[64] H. C. Spruit, *Differential rotation and magnetic fields in stellar interiors*, Astron. Astrophys. **349**, 189

(1999), https://ui.adsabs.harvard.edu/abs/1999A%26A... 349..189S/abstract.

[65] H. C. Spruit, *Dynamo action by differential rotation in a stably stratified stellar interior*, Astron. Astrophys. **381**, 923 (2002).

[66] J. Fuller and L. Ma, *Most black holes are born very slowly rotating*, Astrophys. J. **881**, L1 (2019).

[67] M. Zaldarriaga, D. Kushnir, and J. A. Kollmeier, *The expected spins of gravitational wave sources with isolated field binary progenitors*, Mon. Not. R. Astron. Soc. **473**, 4174 (2018).

[68] S. S. Bavera, M. Fishbach, M. Zevin, E. Zapartas, and T. Fragos, *The $\chi_{\rm eff}$—z correlation of field binary black hole mergers and how 3G gravitational-wave detectors can constrain it*, Astron. Astrophys. **665**, A59 (2022).

[69] C. Talbot and E. Thrane, *Determining the population properties of spinning black holes*, Phys. Rev. D **96**, 023012 (2017).

[70] Sometimes *singular* Beta distributions are also allowed. Singular Beta distributions give $p(\chi) \to \infty$ as $\chi \to 0$, which is also precluded in Fig. 8.

[71] J. Fuller, D. Lecoanet, M. Cantiello, and B. Brown, *Angular momentum transport via internal gravity waves in evolving stars*, Astrophys. J. **796**, 17 (2014).

[72] J. Fuller, M. Cantiello, D. Lecoanet, and E. Quataert, *The spin rate of pre-collapse stellar cores: Wave-driven angular momentum transport in massive stars*, Astrophys. J. **810**, 101 (2015).

[73] C. Chan, B. Müller, and A. Heger, *The impact of fallback on the compact remnants and chemical yields of core-collapse supernovae*, Mon. Not. R. Astron. Soc. **495**, 3751 (2020).

[74] H.-T. Janka, A. Wongwathanarat, and M. Kramer, *Supernova fallback as origin of neutron star spins and spin-kick alignment*, Astrophys. J. **926**, 9 (2022).

[75] S. Vitale, S. Biscoveanu, and C. Talbot, *Spin it as you like: The (lack of a) measurement of the spin tilt distribution with LIGO-Virgo-KAGRA binary black holes*, Astron. Astrophys. **668**, L2 (2022).

[76] T. Damour, *Coalescence of two spinning black holes: An effective one-body approach*, Phys. Rev. D **64**, 124013 (2001).

[77] E. Racine, *Analysis of spin precession in binary black hole systems including quadrupole-monopole interaction*, Phys. Rev. D **78**, 044021 (2008).

[78] P. Schmidt, F. Ohme, and M. Hannam, *Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter*, Phys. Rev. D **91**, 024043 (2015).

[79] At the same time, the *mean* $\chi_{\rm eff}$ among the binary population, found to be $\mu = 0.04^{+0.05}_{-0.09}$, remains consistent with zero.

[80] For example, the coordinate transformation from component spins to $\chi_{\rm p}$ forces $p(\chi_{\rm p})$ to zero at $\chi_{\rm p} = 0$.

[81] https://github.com/tcallister/autoregressive-bbh-inference/.

[82] Thomas A. Callister and Will M. Farr, *A parameter-free tour of the binary black hole population*, Zenodo (2023), 10.5281/zenodo.7600140.

[83] https://www.gw-openscience.org.

[84] Astropy Collaboration, *Astropy: A community* PYTHON *package for astronomy*, Astron. Astrophys. **558,** A33 (2013).

[85] Astropy Collaboration, *The Astropy Project: Building an open-science project and status of the v2.0 core package*, Astron. J. **156,** 123 (2018).

[86] A. Collette, T. Kluyver, T. A. Caswell, J. Tocknell, J. Kieffer *et al.*, h5py: 3.2.1 (2021).

[87] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Comput. Sci. Eng. **9,** 90 (2007).

[88] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen *et al.*, *Array programming with* NUMPY, Nature (London) **585,** 357 (2020).

[89] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy *et al.*, SCIPY 1.0: *Fundamental algorithms for scientific computing in* PYTHON, Nat. Methods **17,** 261 (2020).

[90] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GW190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object*, Astrophys. J. **896,** L44 (2020).

[91] https://www.gw-openscience.org.

[92] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *The LIGO Open Science Center*, J. Phys. Conf. Ser. **610,** 012021 (2015).

[93] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo*, SoftwareX **13,** 100658 (2021).

[94] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GWTC-1: A Gravitational-Wave Transient Catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs*, Phys. Rev. X **9,** 031040 (2019).

[95] Available at https://dcc.ligo.org/LIGO-P1800370/public.

[96] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run*, Phys. Rev. X **11,** 021053 (2021).

[97] Available at https://dcc.ligo.org/LIGO-P2000223/public.

[98] Available at https://zenodo.org/record/5546663.

[99] https://zenodo.org/record/5636816.

[100] C. Talbot and J. Golomb, *Growing pains: Understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy*, Mon. Not. R. Astron. Soc. **526,** 3495 (2023).

[101] R. Essick and W. Farr, *Precision requirements for Monte Carlo sums within hierarchical Bayesian inference*, arXiv:2204.00461.

[102] S. Miller, T. A. Callister, and W. M. Farr, *The low effective spin of binary black holes and implications for individual gravitational-wave events*, Astrophys. J. **895,** 128 (2020).