Probing correlations in the binary black hole population with flexible models

Jack Heinzel[1,2,*] Sylvia Biscoveanu,[1,2,3,†] and Salvatore Vitale[1,2]

[1]*LIGO, Massachusetts Institute of Technology, 77 Massachusetts Avenue,*
Cambridge, Massachusetts 02139, USA
[2]*Kavli Institute for Astrophysics and Space Research and Department of Physics,*
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
[3]*Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA),*
Northwestern University, 1800 Sherman Avenue, Evanston, Illinois 60201, USA

The astrophysical formation channels of binary black hole systems predict correlations between their mass, spin, and redshift distributions, which can be probed with gravitational-wave observations. Population-level analysis of the latest LIGO-Virgo-KAGRA catalog of binary black hole mergers has identified evidence for such correlations assuming linear evolution of the mean and width of the effective spin distribution as a function of the binary mass ratio and merger redshift. However, the complex astrophysical processes at play in compact binary formation do not necessarily predict linear relationships between the distributions of these parameters. In this work, we relax the assumption of linearity and instead search for correlations using a more flexible cubic spline model. Our results suggest a nonlinear correlation between the width of the effective spin distribution and redshift. We also show that the LIGO-Virgo-Kagra collaborations may find convincing Bayesian evidence for nonlinear correlations by the end of the fourth observing run, O4. This highlights the valuable role of flexible models in population analyses of compact-object binaries in the era of growing catalogs.

I. INTRODUCTION

The growing catalog of gravitational-wave observations of binary black hole (BBH) mergers has allowed for increasingly detailed probes of the population properties of these systems, with the ultimate goal of revealing how they form and evolve. Analysis of the third gravitational-wave transient catalog (GWTC-3) [1] of the LIGO-Virgo-Kagra collaboration (LVK) [2–6] found evidence for substructure in the BBH primary mass distribution [7–9] beyond a single power law [10] plus Gaussian peak [11] and for preferentially equal-mass mergers [12,13]. The black hole spin distribution favors small [14] but likely nonzero spins [15–20], and the distribution of the spin tilts relative to the orbital angular momentum is consistent with isotropy, while the distribution of effective aligned spins indicates a slight preference towards positively aligned spins [7,20–23]. The merger rate evolves with redshift at a rate consistent with the cosmic star formation rate [24]. Taken altogether, these constraints imply that there are probably multiple formation channels [25–27] shaping the BBH population, although no definitive evidence for subpopulations with different properties has been identified [28–30].

While most previous analyses of the BBH population have relied on phenomenological population models with simple parametric functional forms, recent work has explored the use of "nonparametric" models like splines [23,31], Dirichlet processes [32], generic mixture models [8,33,34], binned Gaussian processes [7,35,36], and auto-regressive models [37]. Despite including many more free parameters than the phenomenological models, these nonparametric models offer increased flexibility to fit the data without imposing a specific functional form. This avoids the issue of model misspecification, e.g., [38], at the cost of a clear mapping between features observed in the data and those predicted by astrophysical theory.

As the observed population of BBH mergers grows, population analyses have moved beyond modeling the mass, spin, and redshift distributions independently towards searching for correlations between these parameters. Such correlations are expected both within individual formation channels and due to the superposition of subpopulations forming via distinct channels [39–48]. For example, BBHs formed via isolated binary evolution may exhibit correlations due to the relationship between metallicity and the efficiency of angular momentum transport via stellar winds. These winds remove mass and spin down the progenitor, meaning that more massive systems would prefer higher spins. They would also be more common at high redshift, where stellar winds are less efficient because of lower

[*]heinzelj@mit.edu
[†]NASA Einstein Fellow

metallicities [49]. A mass-spin correlation is also expected for systems formed dynamically via hierarchical mergers in dense environments, since remnants of previous mergers that go on to merge again will be both more massive and rapidly spinning, e.g., [50–52].

Evidence for a correlation between BBH mass and spin was obtained using a phenomenological model where the mean and width of the distribution of effective aligned spin (χ_{eff}) vary linearly as a function of the mass ratio, such that systems with more unequal mass ratios have larger effective aligned spins [7,53]. An approach using copula density functions that ensure fixed marginal distributions in the presence of a correlation identifies this correlation at similar significance [54,55]. Weaker evidence using the linear correlation model also hints at a potential correlation between the effective spin distribution and the total mass and primary mass [56–58], although a broadening of the spin distribution at the highest masses can be explained due to the relative dearth of observations in this regime, leading to a more uncertain measurement [34]. A likely correlation between the width of the effective spin distribution and redshift has also been identified [58]. Recently, using a nonparametric method, Ref. [59] reports a correlation in primary mass-redshift space, arising from two subpopulations. However, some results of this work are in tension with previous population analyses, including both parametric and nonparametric methods, e.g., [31,34,37,60].

In this work, we build upon previous methods looking for correlations between individual pairs of parameters describing the BBH mass, spin, and redshift distributions but adopt a more flexible model for the shape of the correlation. Specifically, we assume the BBH effective spin distribution is described by a Gaussian with an unknown mean and width, both of which may correlate with either the mass or redshift such that the shape of the correlation is given by a cubic spline. Our model recovers evidence for the previously identified correlations between effective spin and mass ratio and redshift but prefers a nonlinear shape for the spin-redshift correlation. This result highlights the important role that flexible population models will play in identifying model misspecification as the catalog of BBH merger observations grows. The rest of this work is structured as follows.

In Sec. II we briefly describe our statistical assumptions, which are standard in gravitational wave (GW) population inference, then describe our model for probing correlations in more detail. In Sec. III we present the constraints on the BBH data collected thus far [1], which tend to be broadly consistent with previous studies apart from some evidence for nonlinearity in the $\chi_{\text{eff}} - z$ correlation. In Sec. IV we make projections for the future: How well can nonlinearity be measured with future observations? In a pair of simulated universes with nonlinear correlations in $\chi_{\text{eff}} - q$ and $\chi_{\text{eff}} - z$, nonlinearity does not reveal itself in the $\chi_{\text{eff}} - q$ distribution with 400 detections, but there is strong evidence for nonlinearity in the $\chi_{\text{eff}} - z$ distribution

with 400 detections. While the detectability of nonlinearity ultimately is subject to how nonlinear the true correlated distribution is, we show it is possible to detect nonlinearity in the near future. Finally, we conclude in Sec. V.

II. PROBING CORRELATIONS: PRIORS AND PARAMETRIZATION

The goal of population modeling is to infer the distribution from which an ensemble of observations is drawn. This can be accomplished using hierarchical Bayesian inference, which takes a multistage approach by first characterizing individual observations and then combining them on a population level. In a GW context, these individual observations are noisy, so the statistical likelihood of observing the data given a population model $p(\theta|\Lambda)$ parametrized by Λ must be marginalized over the possible GW parameters [61]:

$$\mathcal{L}(d|\Lambda) = \int d\theta \mathcal{L}(d|\theta) p(\theta|\Lambda). \tag{1}$$

Here, d represents the detected data, and θ represents the unknown source parameters, like the binary masses, spins, and redshift.

Furthermore, BBHs suffer a Malmquist bias; they are not all equally detectable. To account for this bias, we must define the detection efficiency,

$$\alpha(\Lambda) = \int_{\mathcal{D}_\uparrow} dd \int d\theta \mathcal{L}(d|\theta) p(\theta|\Lambda), \tag{2}$$

which is the fraction of events in the population $p(\theta|\Lambda)$ which probabilistically generate detectable data ($d \in \mathcal{D}_\uparrow$). Assuming the events are distributed in time by a Poisson process and marginalizing over the Poisson rate parameter with a uniform-in-log prior, we obtain the rate-marginalized hierarchical likelihood [35,61–65]. In particular, with a collection of N_{obs} events with data $\{d_i\}_{i=1}^{N_{\text{obs}}}$ passing a detection threshold

$$\mathcal{L}(\{d_i\}_{i=1}^{N_{\text{obs}}}|\Lambda) \propto \prod_{i=1}^{N_{\text{obs}}} \frac{\mathcal{L}(d_i|\Lambda)}{\alpha(\Lambda)}. \tag{3}$$

We use this likelihood to sample from the hyperparameters Λ. In practice, we estimate the marginal integrals in Eqs. (1) and (2) with Monte Carlo estimators; see Refs. [66–68] for details.

A common approach tackles the hierarchical inference problem by phenomenologically parametrizing the unknown source distribution as, e.g., power laws, Gaussians, etc., and inferring these hyperparameters given the data observations. Previous studies have identified simple and successful parametrization schemes; we choose the astrophysically motivated parametrizations POWER-LAW

+PEAK [11] and POWER-LAW REDSHIFT [69] for the mass $p(m_1, q|\Lambda)$ and redshift $p(z|\Lambda)$ distributions, respectively. The primary mass distribution is parametrized as a smoothed power law plus a Gaussian component, and a smoothed power law for the mass ratio (hyperparameters are the minimum and maximum BH mass, power-law index, low mass smoothing parameter, mean and width of the Gaussian, fraction of BBHs in the Gaussian component, and the power-law index for the mass ratio). The redshift distribution is modeled as a power law with a single power-law index hyperparameter.

For the spins, we project the six dimensional spin distribution to a one dimensional parameter $\chi_{\rm eff}$ describing the leading order spin effect on the inspiral evolution of the binary [70–72], which is often the most well measured spin parameter [73,74]. We then model the $\chi_{\rm eff}$ population as a truncated Gaussian distribution with a variable mean and width [21,75]:

$$p(\chi_{\rm eff}|\theta, \Lambda) = \frac{1}{\sigma(\theta, \Lambda)} \frac{\phi\left(\frac{\chi_{\rm eff} - \mu(\theta, \Lambda)}{\sigma(\theta, \Lambda)}\right)}{\Phi\left(\frac{1 - \mu(\theta, \Lambda)}{\sigma(\theta, \Lambda)}\right) - \Phi\left(\frac{-1 - \mu(\theta, \Lambda)}{\sigma(\theta, \Lambda)}\right)}, \quad (4)$$

where the truncation restricts the support of the distribution to be $[-1, 1]$, which reflects the physical constraint that the magnitude of $\chi_{\rm eff}$ cannot exceed 1 for BH spins bounded by the Kerr limit. ϕ is the standard Gaussian, and Φ is the standard error function,

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \qquad \Phi(x) = \int_{-\infty}^{x} \phi(s) ds. \quad (5)$$

This is the approach used by Refs. [7,53,56,58] to explore models where $\mu(\theta)$ and $\sigma(\theta)$ are linear functions of primary mass ($\theta = m_1$), mass ratio ($\theta = q$), and redshift ($\theta = z$),

respectively. Building on this previous work, we model $\mu(\theta)$ and $\sigma(\theta)$ with spline functions.

Spline functions are becoming increasingly popular in GW population inference, primarily because they are innately flexible and fast to evaluate, but also because they are easily parametrized by their nodes. While splines do not assume much structure, they do fail to probe structure below the scale of the node separation length. For this reason, one should include the number of nodes as a model hyperparameter or repeat the analysis varying the choice of the number of nodes.

In this work, we use a cubic spline model, where nodes are interpolated using cubic polynomials which preserve continuity in the function and first and second derivatives at the nodes ($\mathcal{C}^2$ functions). Given node locations, this provides all but two conditions to set the four coefficients for each cubic polynomial; the final two conditions are given at the end points, typically imposed by setting the second derivative to zero. Defined in this way, the node positions fully determine the spline curve. Splines also have the advantage of approximate locality, meaning they can fit a structure in one region of parameter space independently of the behavior of the spline far away (separated by many nodes) [76].

For all the inferences we present below, we use priors shown in Table I. We use the OVERALL samples for GWTC1 events [77], PRECESSINGSPINIMRHM for the events first identified in GWTC2 [78] and the C01: IMRPHENOMXPHM samples for the GWTC2.1 and GWTC3 events [79,80], and the search sensitivity estimates provided in Ref. [81]. We use the GWPOPULATION package for constructing the hierarchical likelihood [82], and compute Bayesian evidences while sampling the hyperposterior using the DYNESTY implementation in BILBY [83,84]. To efficiently evaluate spline functions, we use the CACHED_INTERPOLATE package introduced in Ref. [23]. In addition, because

TABLE I. Priors for each hyperparameter used in our model. The upper panel describes the standard priors, except where they are reduced to improve the sampling efficiency without cutting off the hyperposterior ($m_{\rm max}$ and λ). The lower panel describes the prior assumed for our spline parameters.

Hyperparameter	Description	Prior
α	m_1 power-law index	$U(-4, 12)$
β	q power-law index	$U(-4, 7)$
$m_{\rm max}$	Maximum BH mass	$U(60M_\odot, 100M_\odot)$
$m_{\rm min}$	Minimum BH mass	$U(2M_\odot, 10M_\odot)$
δ_m	Low-mass smoothing parameter	$U(0M_\odot, 10M_\odot)$
μ_m	m_1 Gaussian component mean	$U(20M_\odot, 50M_\odot)$
σ_m	m_1 Gaussian component standard deviation	$U(1M_\odot, 10M_\odot)$
λ	Fraction of BBHs in Gaussian component	$U(0, 0.2)$
λ_z	z power-law index	$U(-2, 10)$
$\mu_{\chi_{\rm eff}:n}$	nth spline node for the $\chi_{\rm eff}$ mean	$U(-1, 1)$
$\ln \sigma_{\chi_{\rm eff}:n}$	nth spline node for the $\chi_{\rm eff}$ standard deviation	$U(-5, 0)$

we estimate the population likelihood [Eq. (3)] with Monte Carlo integrals, there is inherent uncertainty associated with each likelihood estimate. To avoid biased inference, we cut hyperposterior samples with log-likelihood uncertainty greater than 1, following the recommendation of Ref. [68].

III. CORRELATIONS IN GWTC-3

Using the catalog of 69 BBHs described in Ref. [7] passing a detection-pipeline-computed false alarm rate threshold of 1 yr^{-1}, we search for evidence of nonlinear correlations between the spin distribution and the mass ratio, redshift, and the source frame primary mass (the latter is described in Appendix A). To do this, we model the χ_{eff} distribution with Eq. (4), and present the results of the individual analyses below.

A. Effective spin distribution and mass ratio

The first correlation we explore is between χ_{eff} and mass ratio q, setting the mean and standard deviations to spline functions of mass ratio. We place the nodes uniformly between $q = 0$ and $q = 1$. Though $q \to 0$ is an unphysical region of parameter space with no observations, the node at $q = 0$ should be thought of as simply a parameter necessary to ensure the model is defined over the entire space, and not an *a priori* statement that BBHs exist here; indeed the model $p(q|m_1, \Lambda)$ is always zero at $q = 0$:

$$\mu(\theta) = S\big(q|(0, \mu_{\chi_{\mathrm{eff}}:0}), \ldots, (1, \mu_{\chi_{\mathrm{eff}}:N})\big)$$

$$\ln \sigma(\theta) = S\big(q|(0, \ln \sigma_{\chi_{\mathrm{eff}}:0}), \ldots, (1, \ln \sigma_{\chi_{\mathrm{eff}}:N})\big), \quad (6)$$

where $S(x|(X_1, Y_1), \ldots, (X_N, Y_N))$ represents the cubic spline function in the variable x passing through the nodes with x coordinates $X_1 < X_2 < \cdots < X_N$ and corresponding y coordinates $Y_1, Y_2, \ldots, Y_N$. In all our models, we fix the x coordinates to reduce the dimension of the inference.

In Fig. 1, we show a comparison between our results with the linear model of Ref. [53] to the spline model with four nodes. While our spline model results are broadly consistent with the linear model, they generically feature broader credible intervals towards extreme mass ratios $q \to 0$. We argue that this is an advantage of the spline model, as the data should have less information about BBHs with unequal mass ratios as they are less common in the detected population, e.g., [12,13], and be completely uninformative about events with mass ratio $q \to 0$. All of the 69 BBHs in the catalog are inconsistent with vanishing mass ratios, and even more, our population model with hyperparameter $m_{\mathrm{min}} \geq 2M_\odot$ (the minimum BH mass must exceed $2M_\odot$, see Ref. [11]) requires zero support at $q \to 0$. Being an approximately local model, the spline model can fit the structure at near-equal mass ratios, while simultaneously saying nothing about the behavior of extreme mass
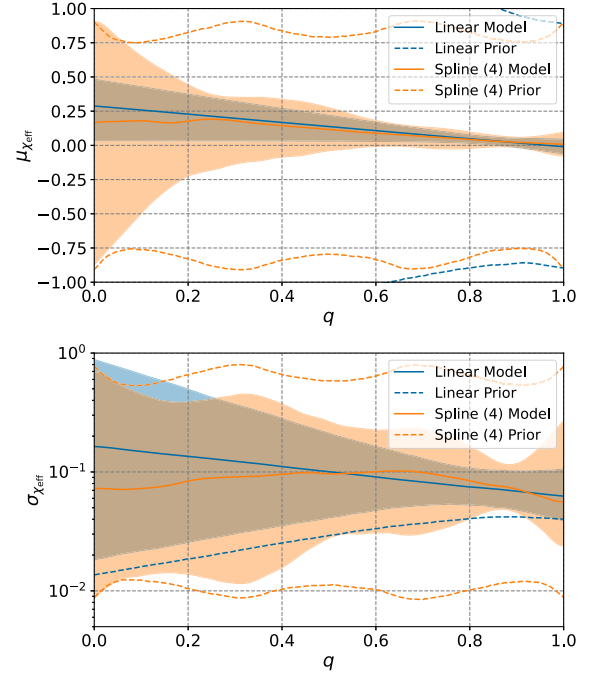

FIG. 1. A comparison between the linear model and four node spline model for correlation between χ_{eff} and mass ratio, inferred using GWTC-3. Solid lines represent the median, while the shaded region represents the central 90% credible interval. Dashed lines show the upper and lower boundaries of the prior 90% interval. The upper panel shows the mean of the χ_{eff} Gaussian as a function of mass ratio q, while the lower panel is the standard deviation of the χ_{eff} distribution. For the spline model, the model inference is increasingly prior driven at small mass ratios, where there is little information in the data.

ratio events. Hence, the posterior approaches the prior as $q \to 0$. We also repeat the analysis with 3–6 nodes and obtain results consistent with the four-node analysis presented here, see Appendix B.

To compute the significance of any evolution with mass ratio, we compute the derivative of the inferred evolution with respect to mass ratio. Spline functions are easily differentiable, and so we can compute the slope of the spline function at an arbitrary point q^*. We choose the fiducial value of $q^* = 0.9$, as this appears to be a well constrained region and a good proxy for understanding the evolution of the spin population in the region of near-equal mass ratios. This gives us posteriors on the slope at this point q^* for each model, which we show in Fig. 2. We then compute the significance of an evolution in the mean of the Gaussian as a function of mass ratio from the fraction of the posterior support with positive/negative slope. The linear model has a negative slope with significance ~98%, as initially reported in Ref. [53]. The spline models have negative slope with more inconclusive significance: ~70%–80%. We show all calculated slope significances at their fiducial values in Table III in Appendix B.
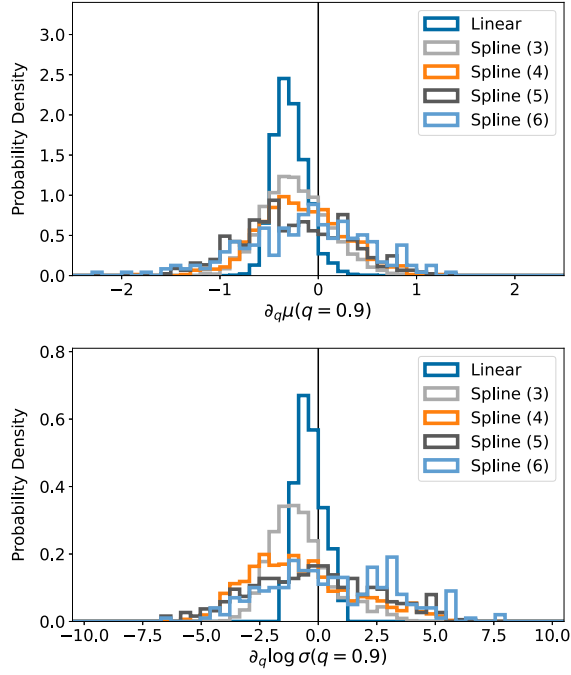
FIG. 2.   Posteriors on the derivative of the mean and width of the $\chi_{\rm eff}$ distribution as a function of mass ratio, calculated at $q^* = 0.9$, and inferred using GWTC 3. The linear model concludes that the mean of the $\chi_{\rm eff}$ distribution decreases as mass ratio approaches 1, with a significance of $\sim 98\%$. The spline models are more agnostic, finding significances of $\sim 60\%$–75%.

We note that this is not the only statistic one could use to quantify the significance. For example, we could consider the difference over widely separated regions in parameter space, e.g., the proportion of samples with mean $\mu(q = 1) < \mu(q = 0.4)$. These statistics tend to find a similar level of significance.

B. Effective spin distribution and redshift

Next, we turn our attention to the correlation between $\chi_{\rm eff}$ and redshift z. Reference [58] examined a correlation between redshift and the effective spin distribution, and discovered evidence for a broadening in the effective spin distribution, a positive correlation between the width $\sigma(z)$ of the $\chi_{\rm eff}$ Gaussian and the redshift, and no evidence for any trend in the mean of the Gaussian. In their analysis, Ref. [58] parametrized $\mu(z) = \mu_0 + \delta\mu_z(z - 0.5)$ and $\log_{10}\sigma(z) = \log\sigma_0 + \delta\log\sigma_z(z - 0.5)$ as linear models and quantified the significance of the measured broadening using the posterior on the slope $\delta\log\sigma_z$.

In a similar approach, we model the $\chi_{\rm eff}$ correlation with redshift using Eq. (4), only we use

$$\mu(\theta) = S\big(z|(0, \mu_{\chi_{\rm eff}:0}), \ldots, (2.3, \mu_{\chi_{\rm eff}:N})\big)$$
$$\ln\sigma(\theta) = S\big(z|(0, \ln\sigma_{\chi_{\rm eff}:0}), \ldots, (2.3, \ln\sigma_{\chi_{\rm eff}:N})\big). \quad (7)$$
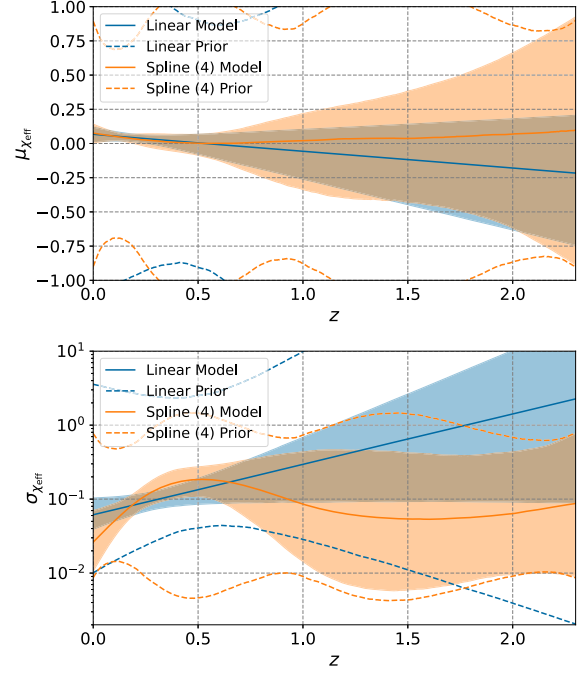

FIG. 3.   A comparison between the linear model and four node spline model for correlation between $\chi_{\rm eff}$ and redshift, inferred using GWTC-3. Solid lines represent the median, while the shaded region represents the central 90% credible interval, and the dashed lines show the boundary of the prior 90% interval. The upper panel shows the mean of the $\chi_{\rm eff}$ Gaussian as a function of redshift z, while the lower panel represents the standard deviation of the $\chi_{\rm eff}$ distribution. We emphasize that there is very little information in the data at $z \gtrsim 1$, hence the spline inference is prior driven at these higher redshifts. This is in contrast to the linear model [58].

In our model, the first and last nodes are placed at redshifts $z = 0$ and $z = 2.3$, the maximum redshift we assume in the POWER-LAW REDSHIFT model [85]. The hyperparameters associated to the splines are the y coordinates of the nodes. We explored models that fix the nodes uniformly between the first and last nodes; however, these resulted in nodes too coarsely spread at small redshift to optimally fit the structure. Additionally, there is limited information in the data thus far to constrain the population much beyond redshift $z \gtrsim 1$, so a better node spacing should place nodes tightly at small redshift and more loosely at high redshift. Heuristically, we found that a linear spacing in $z^{1/2}$ places nodes satisfactorily.

Using these models to infer the BBH population hyperparameters given the GWTC-3 dataset, we infer the evolution of the mean and width of the $\chi_{\rm eff}$ Gaussian as a function of redshift. We show a comparison between a linear model and the spline model with four nodes in Fig. 3. We show the results for a collection of analyses with 3–6 spline nodes in Appendix B in Fig. 10 and the associated model evidences in Fig. 9.

First, we quantify the confidence of a broadening slope for each model. To this end, we compute the slope of the width of the Gaussian at a fiducial value z^*, $\partial \log \sigma(z)z = z^*$ for all models. We choose the fiducial value $z^* = 0.2$ (see Appendix C for a discussion on this choice), and show histograms of the slopes for each model in Appendix C in Fig. 11. We can then quantify the significance of the increase by the proportion of the posterior with slope greater than zero.

There are varying degrees of evidence that the $\chi_{\rm eff}$ distribution is broadening as a function of redshift at $z^* = 0.2$, depending on the model used. The mean is decreasing at $\sim80\%$–95% confidence (the posterior support with slope less than 0), depending on the model assumed. The width is increasing at 90%–98.6% confidence, consistent with the finding of Ref. [58] that the width of the $\chi_{\rm eff}$ distribution broadens with increasing redshift. We collect all significances calculated at fiducial values in Table III in Appendix B.

It also appears that there is some nonlinearity in the evolution of the width as a function of redshift. To

understand if this degree of inferred nonlinearity is expected in a universe with a linear correlation, we perform the spline model inference with four nodes on 12 catalogs of 69 events drawn from a linearly correlated universe (see Appendix C for details on the selection procedure and the generation of the synthetic catalogs). We then compute the derivative at the first three nodes z_0, z_1, and z_2 (ignoring the last node since the posterior is essentially the prior there) and calculate the Jensen-Shannon (JS) divergence [86,87] between the slope posteriors. A linearly correlated universe would theoretically have slope posteriors all consistent with the true value, however there will be some random scatter between the posteriors, measured by the JS divergence between them. Looking at a scatter plot of the divergence between the slope at z_0 and z_1 on the x axis and between z_1 and z_2 on the y axis in Fig. 4, notice that there are no divergences from a linearly correlated universe which is more extreme than the GWTC-3 divergences in both axes. This points towards a nonlinear trend in the width of the $\chi_{\rm eff}$ distribution as a function of redshift,
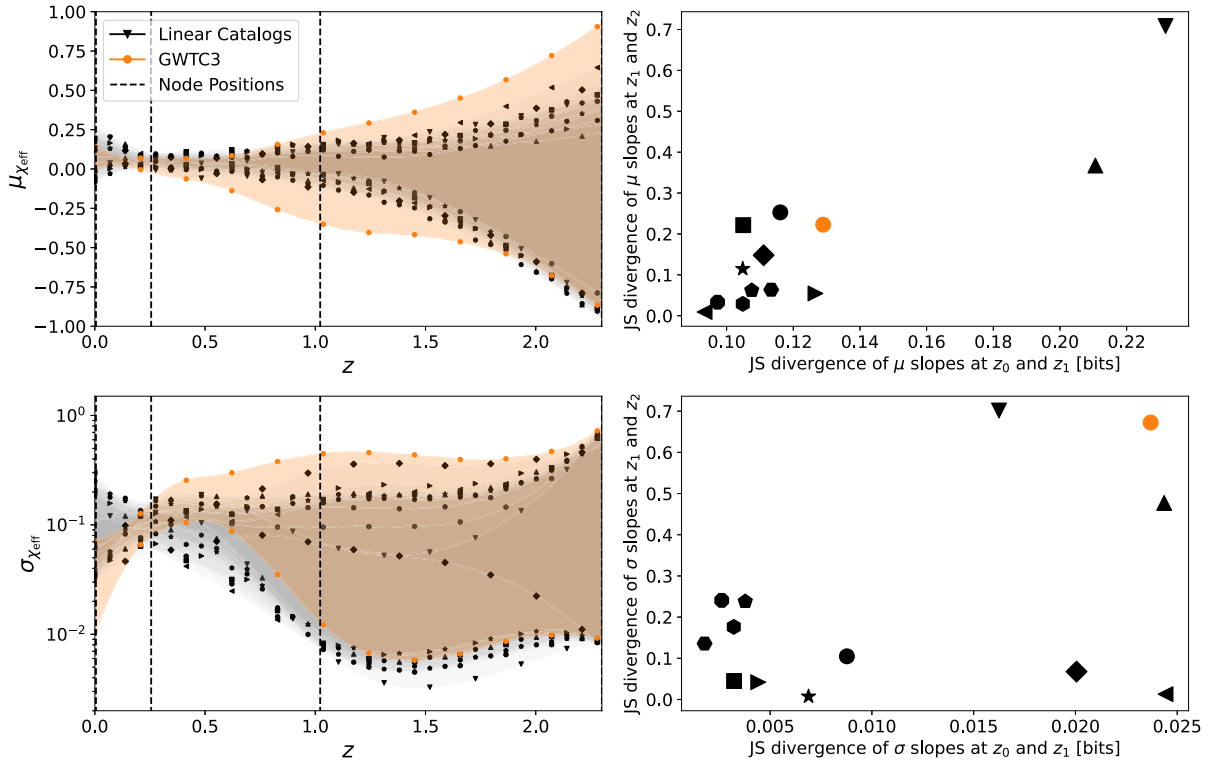

FIG. 4.   Comparison between the spline inference on a linearly correlated universe and the inference on GWTC-3. The first row shows the mean and the second row the standard deviation. The left column shows the 90% credible intervals inferred using GWTC-3 (orange) and the 12 synthetic catalogs of 69 events drawn from a linearly correlated universe (gray). The right column shows the Jensen-Shannon (JS) divergences (in bits) between slope posteriors taken at the first two nodes on the x axis, and between the second and third nodes on the y axis. For each linear catalog there is a black point on the scatter plot, and the orange point corresponds to the GWTC-3 divergences. This represents the posterior difference between the two slopes, quantifying the nonlinearity. The further the point is from the origin, the more confident we are in the inferred nonlinearity. Notice there is no inference on a linear universe which is more nonlinear than GWTC-3 in both dimensions.

though the Bayesian evidence [$\mathcal{Z} = \int \pi(\Lambda)\mathcal{L}(\{d_i\}|\Lambda)d\Lambda$] is not yet conclusive (see Fig. 9 in Appendix B).

IV. FUTURE PROSPECTS: CAN WE DETECT NONLINEARITY IN O4?

Flexible spline models stand in contrast to linear models, where the correlation around each spline node is independently inferred, rather than enforcing a consistent slope across the whole parameter space. At the conclusion of O3, these flexible models highlight our lack of knowledge about poorly constrained regions. While we have some hints of a nonlinear correlation in $\chi_{\mathrm{eff}} - z$, there is not yet a definitive preference in the Bayesian evidence for a non-linear correlation. This may change by the end of the fourth observing run, O4.

To predict how well we may actually detect nonlinear correlations in the future, we produced two synthetic catalogs of 200 events and 400 events for two different possible universes (so four catalogs total). The events are detected by a network of LIGO interferometers (located at Hanford and Livingston), assuming the fiducial O4 noise spectra with an average binary neutron star merger inspiral range of 160 Mpc [88]. We draw GW events from a few different populations with true hyperparameters given in Table II. These populations are consistent with the data collected by the LVK thus far, analyzed with the models we presented above. We use the waveform model IMRPHENOMXP with PRECVERSION=104 [89], and use the heterodyning/relative binning scheme of Refs. [90–92] to efficiently sample the GW event parameter posteriors.

We then select detected events on a network matched-filter SNR $\rho_{\mathrm{mf,HL}} > 9$, and generate a set of Monte Carlo injections for estimating the selection efficiency consistent with this detection criterion [93]. This kind of selection criterion implicitly depends on the true binary system parameters, which is not a physical assumption [94]. This means that the hierarchical likelihood of Eq. (3) should be slightly modified to match this different selection criterion. However, we use the same likelihood of Eq. (3) as an approximation. The bias is likely small for catalogs of relatively small size, such as the ones in this work, but it should be noted that this may be present in our inferences on synthetic catalogs [65].

We recover each simulated catalog assuming both a linear model and the spline model with four nodes. We restrict to one spline analysis using four nodes to limit the computational expense.

We decided to explore two node placement options. First, we recover using our original node placement scheme, placed in the same manner as described above (linear in q and $z^{1/2}$). This node placement does not match the true node placement (Table II), and the inference finds more "bumps" in the correlation than are truly there. To this end, we also recover using the true node placement scheme. See Appendix D for a discussion on node placement and bias in the inference.

We show our results comparing the linear model and the spline model with our original node placement (linear in q and $z^{1/2}$) in Fig. 5.

In a catalog of 200 detections, the model evidences indicate no significant preference for a linear model or the

TABLE II. Hyperparameters for the simulated universes with nonlinear correlation. In the left column are the hyperparameters for the universe with a correlation between χ_{eff} and mass ratio, and in the right column is the correlation with redshift. The correlation functions are themselves splines, with node placements given by, e.g., the $(z_N, \mu_{\chi_{\mathrm{eff}}:N})$ coordinate pairs.

| Hyperparameter	Description	$q - \chi_{\mathrm{eff}}$ correlation	$z - \chi_{\mathrm{eff}}$ correlation
α	m_1 power-law index	3	3
β	q power-law index	1	1
$m_{\max}$	Maximum BH mass	$85 M_\odot$	$85 M_\odot$
$m_{\min}$	Minimum BH mass	$5 M_\odot$	$5 M_\odot$
δ_m	Low-mass smoothing parameter	$3 M_\odot$	$3 M_\odot$
μ_m	m_1 Gaussian component mean	$35 M_\odot$	$35 M_\odot$
σ_m	m_1 Gaussian component standard deviation	$5 M_\odot$	$5 M_\odot$
λ	Fraction of BBHs in Gaussian component	0.03	0.03
λ_z	z power-law index	2	2
$(x_0, \mu_{\chi_{\mathrm{eff}}:0})$	First mean spline node coordinates	(0,0.4)	(0,0)
$(x_1, \mu_{\chi_{\mathrm{eff}}:1})$	Second mean spline node coordinates	(0.4,0.3)	(0.3,0)
$(x_2, \mu_{\chi_{\mathrm{eff}}:2})$	Third mean spline node coordinates	(0.8,0.05)	(0.65,0)
$(x_3, \mu_{\chi_{\mathrm{eff}}:3})$	Fourth mean spline node coordinates	(1,0.02)	(2.3,0)
$(x_0, \ln \sigma_{\chi_{\mathrm{eff}}:0})$	First standard deviation spline node coordinates	$(0, -2.5)$	$(0, -3.5)$
$(x_1, \ln \sigma_{\chi_{\mathrm{eff}}:1})$	Second standard deviation spline node coordinates	$(0.4, -2.5)$	$(0.3, -2)$
$(x_2, \ln \sigma_{\chi_{\mathrm{eff}}:2})$	Third standard deviation spline node coordinates	$(0.8, -2.5)$	$(0.65, -1.5)$
$(x_3, \ln \sigma_{\chi_{\mathrm{eff}}:3})$	Fourth standard deviation spline node coordinates	$(1, -2.5)$	$(2.3, -1.25)$
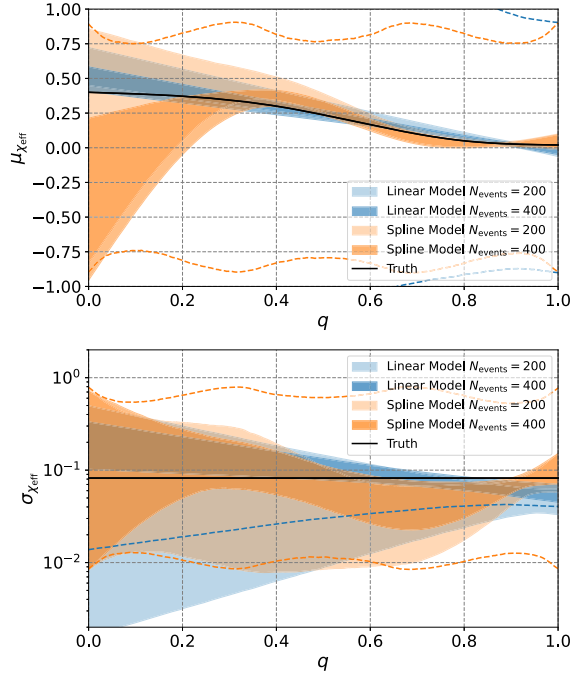
FIG. 5. Inferred mean and width of the χ_{eff} distribution as a function of mass ratio q on two mock catalogs. In blue are the 90% credible intervals when assuming a linear model for the correlation, while in orange are the 90% intervals inferred under the spline model with four nodes. The lighter shades represent the inference on a catalog with 200 events, while the darker shade is with 400 events. Dashed lines represent the prior 90% region, and in solid black is the true correlation.

FIG. 6. Inferred mean and width of the χ_{eff} distribution as a function of redshift z on two mock catalogs. In blue are the 90% credible intervals when assuming a linear model for the correlation, while in orange are the 90% intervals inferred under the spline model with four nodes. The lighter shades are the inference using a catalog of 200 events, while the darker shade is with 400 events. Dashed lines represent the prior 90% region, and in solid black is the true correlation.

spline model. Indeed, the Bayes factor of the spline model over the linear model with 200 events is $\log_{10}\mathcal{B}_{s|l}(200) = 0.00 \pm 0.13$. However, when we increase the number of detections to 400, the data begins to indicate a slight preference for the spline model, although nothing yet conclusive; $\log_{10}\mathcal{B}_{s|l}(400) = 0.28 \pm 0.14$. We emphasize that this holds for a particular choice of a nonlinear correlation. In truth, the correlation may be more or less linear than the one we simulated, which would make evidence of nonlinearity correspondingly more or less definitive at these numbers of events.

To understand how we may constrain the correlation between the spin population and the redshift in the future, we also simulated a universe with a nonlinear correlation between the width of the χ_{eff} distribution and redshift (see Table II), where the model hyperparameters chosen are consistent with the catalog through the end of the third observing run.

We once again produced mock catalogs with 200 and 400 detections, and we show the results of the inferences in Fig. 6. This time, even a false node placement scheme does a good job at fitting the correlation. Furthermore, model evidences become conclusively in favor of the nonlinear model with a catalog of 400 detections. In particular, for the redshift correlation $\log_{10}\mathcal{B}_{s|l}(200) = 1.06 \pm 0.14$
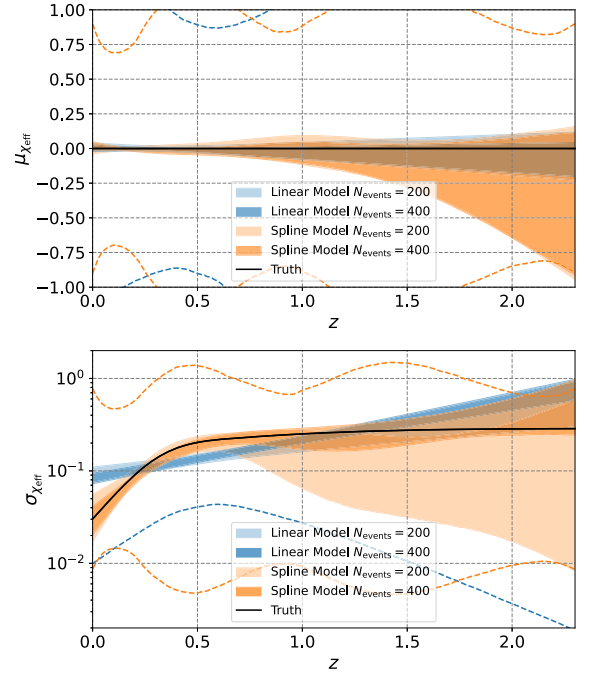
and $\log_{10}\mathcal{B}_{s|l}(400) = 4.09 \pm 0.16$. We also checked that a spline model with nodes placed in the correct locations produce consistent results, as expected.

We reiterate that the results of this projection study depend on the assumed true correlation. However, since the correlation we chose is consistent with the data collected thus far, this demonstrates that we *can* detect nonlinearity in the spin correlation with a catalog of ∼400 detections.

V. CONCLUSIONS

In this paper we presented a flexible model for understanding the correlation between the spin population in χ_{eff} and primary mass m_1, mass ratio q or redshift z. On the LVK data published thus far, we obtain results broadly consistent with previous analyses [7,53,56–58]. Furthermore, because of their flexibility, these spline models highlight the regions of parameter space that drive a measured correlation, and the regions of parameter space which are more uncertain.

In particular, we find that the mean of the χ_{eff} distribution likely increases with mass ratio, the width likely broadens with redshift, and may also broaden for more massive binaries. Importantly, it is also possible that not all of these claims are true at the same time. Perhaps a correlation with

one parameter may masquerade as a correlation with another parameter when analyzed under the false hypothesis. Reference [58] studied this possibility using linear models which simultaneously fit the correlation with m, q, and z, and found that the data is not yet informative enough to answer this question, though it appears possible that all correlations are real. While we do not consider simultaneous models in this work, we do observe the Bayes factors all appear consistent, suggesting the data is not yet informative enough to pick out any mismodeled correlations.

Each of these potential correlations may prove to be important probes of the astrophysical environments which produce BBHs. The observed anticorrelation between the mean of the χ_{eff} distribution with mass ratio may indicate that the BBHs observed come from binaries which experience mass ratio reversal using an optimistic common envelope (CE) prescription [95], or isolated binaries which either undergo a CE phase with large CE efficiency or stable mass transfer with super-Eddington accretion [47,96]. It is also possible that hierarchical mergers in a dense environment can produce the observed correlation; we would expect hierarchical mergers involving just one second-generation black hole from a previous merger to have more extreme mass ratios and higher spins, e.g., [52]. However, in an environment with isotropic spin symmetry, the χ_{eff} distribution must be centered on zero, and so one would expect a *broadening* of the distribution at small mass ratios, not an increase in the mean. In order to produce the observed correlation, then, the environment must break isotropy symmetry somehow [97], e.g., in an AGN disk [44,97]. Finally, if the observed BBHs do not originate from one channel, but a superposition of multiple, e.g., [25,26], a (linear) correlation may arise from a Simpson's-type paradox, where multiple populations naturally separated in $\chi_{eff} - q$ space are interpreted as a correlation [28]. Especially in this last scenario, a flexible nonlinear model will help shed light on the origin of the $\chi_{eff} - q$ correlation as we move into O4.

The correlation with redshift probes evidence for other kinds of pathways toward merging stellar mass BBHs. This observation is commonly explained by connecting the spin-up of a BH or its progenitor with the delay time to merger. If the spin-up mechanism of BH progenitors is stronger for closer separations, the remnant BBH system will radiate energy to GWs more rapidly, and hence merge earlier. Because BBHs are born at a higher rate at higher redshifts, this results in a positive correlation between spin magnitude and redshift [40,98–100]. One potential mechanism is tidal torques: the spin-up due to tidal torques is amplified if the binary is at a smaller initial separation. If the stellar progenitors retain some angular momentum upon collapse, the BH spins should be correlated with the observed redshift at merger [39,101,102]. This is complicated by the expectation that formation in the field produces nearly aligned spins. In this scenario, then there should be a positive correlation between redshift and the mean of the Gaussian; increasing the width requires larger spins with a more isotropic tilt distribution. If there are strong supernova kicks, however, this can even out the spin tilts and give a χ_{eff} distribution more centered on zero [103–106].

Finally, a correlation between χ_{eff} and the primary mass is most naturally understood as a signature of hierarchical mergers. In this paper, we observe a broadening of the χ_{eff} distribution as the primary mass increases, with varying levels of confidence depending on the model (see Appendix A). This is precisely the expectation of a hierarchical merger picture in an environment endowed with isotropy symmetry [52,107–109]. That said, the most recent studies which directly searched for signatures of hierarchical mergers in the LVK catalog strongly disfavor scenarios where all the LVK BBHs are formed hierarchically [110–112], though they cannot be ruled out as a subpopulation.

To understand how we might probe nonlinear correlations in the future, we also simulated nonlinearly correlated universes consistent with the data collected thus far. In a universe with a nonlinear correlation between χ_{eff} and mass ratio, we found that a linear model may still be appropriate with ∼400 detections, although this is of course heavily dependent on how nonlinear the true correlation is. For a nonlinear correlation in the $z - \chi_{eff}$ plane, we find that a linear model may be significantly disfavored as we approach ∼400 detections. Because the nonlinear correlations we assumed were consistent with LVK data collected thus far, it is possible that linear models for correlation will begin to fail by the end of O4.

However, we also encountered a few drawbacks with spline functions. For one, flexible models are intrinsically high dimensional, and thus sampling from the hyperposterior can become significantly more expensive. Second, spline models are perhaps too good at finding nonlinearities, to the point of confidently finding nonlinear features even when they are not present (see Appendix D). To be confident in a spline feature, one should recover the feature varying the number of nodes or the locations of the nodes, or even directly estimate the false alarm probability as in Ref. [113].

In this paper, we argue that flexible models for correlation offer a complementary but equally important perspective compared to strongly parametrized models. Foremost, spline models can effectively fit a much wider range of potential correlations. We do not *a priori* expect nature to provide us with linear correlations, or even correlations that can be well approximated by a line, e.g., [49–52]. The correlations in nature may be strongly nonlinear, or perhaps form as a result of a superposition of subpopulations, and these can only be observed when analyzed with a sufficiently flexible model. Of course, it is possible the correlations will indeed turn out to be linear, however we can only observe such a phenomenon by allowing for the alternative.

APPENDIX A: EFFECTIVE SPIN DISTRIBUTION AND PRIMARY MASS

We also search for potential correlation in the population between the χ_{eff} distribution and the primary mass m_1. Previous studies have searched for the same correlation (see, e.g., Refs. [56,58]) using a linear model for the mean and width in Eq. (4), and discovered weak evidence for a trend in the width of the χ_{eff} distribution.

Using the same spline approach described above, we place nodes between $m_{1,\text{min}} = 2$ and $m_{1,\text{max}} = 100$, linearly spaced in $m_1^{1/3}$. A uniform node spacing does not easily allow for structure at ~ 10–$30 M_{\odot}$, which we would like to probe. Instead, linearly spaced in $m_1^{1/3}$ cluster nodes towards $m_1 \sim 10$–$30 M_{\odot}$ were found to be satisfactory:

$$\mu(\theta) = S\big(m_1|(2, \mu_{\chi_{\text{eff}}:0}), \ldots, (100, \mu_{\chi_{\text{eff}}:N})\big)$$
$$\ln \sigma(\theta) = S\big(m_1|(2, \ln \sigma_{\chi_{\text{eff}}:0}), \ldots, (100, \ln \sigma_{\chi_{\text{eff}}:N})\big). \quad \text{(A1)}$$

We show a comparison between the spline model with four nodes and the linear model, in Fig. 7. The linear model and spline models appear broadly consistent; indeed the model evidences do not show any preference for one model over another (see Appendix B, Fig. 9. We also show the results for all spline runs in Fig. 10 in Appendix B).

We also compute the slope of the mean and width at the fiducial value $m_1^* = 35 M_{\odot}$, and here we find stronger evidence for a positive slope in some spline models than in the linear model. We chose $m_1^* = 35 M_{\odot}$ to coincide
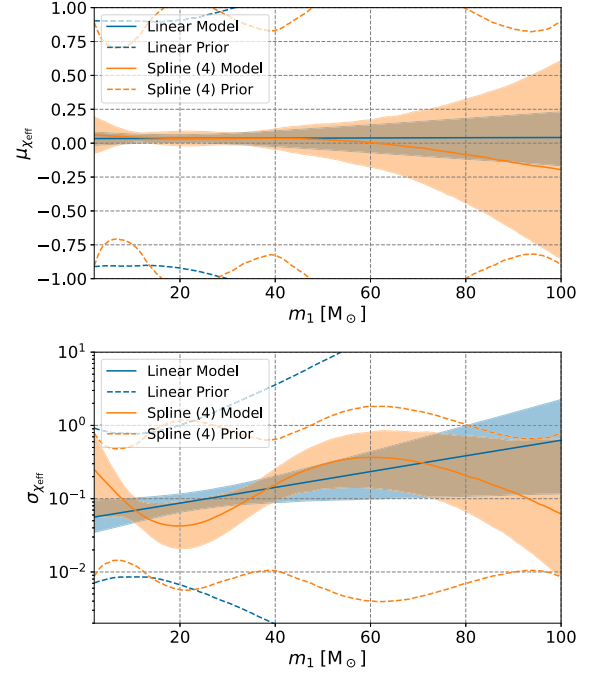

FIG. 7. A comparison between the linear model and four node spline model for correlation between χ_{eff} and primary mass, inferred using GWTC-3. Solid lines represent the median, while the shaded region represents the central 90% credible interval, and the dashed lines show the boundary of the prior 90% interval. The upper panel shows the mean of the χ_{eff} Gaussian as a function of primary mass m_1, while the lower panel represents the standard deviation of the χ_{eff} distribution.

roughly with the location of the Gaussian "peak" in the primary mass distribution [7,11], and so represents a physically interesting region of parameter space. The linear model finds a broadening at higher primary masses with significance $\sim 93\%$, while the spline models vary in significance, the model with four nodes notably exhibits a broadening at 98.7% credence. We show posteriors on the slope at $m_1^* = 35 M_{\odot}$ in Fig. 8.

APPENDIX B: INFERENCES ON GWTC-3, EXTRA ANALYSES AND EVIDENCES

We also run each correlation inference with 3–6 nodes for the mean and standard deviation spline functions. For the $\chi_{\text{eff}} - q$ correlation, we show the constraints on the mean and standard deviation functions in Fig. 9. We do not observe any strong preference for including more nodes in the spline correlation functions; there is a generic "Occam's penalty" for including more nodes beyond what is necessary to appropriately fit the data. We show the evidence of the GWTC-3 data given all our correlation models in Fig. 9; note the evidences are still inconclusive at this stage, though a correlation with mass ratio is somewhat preferred.
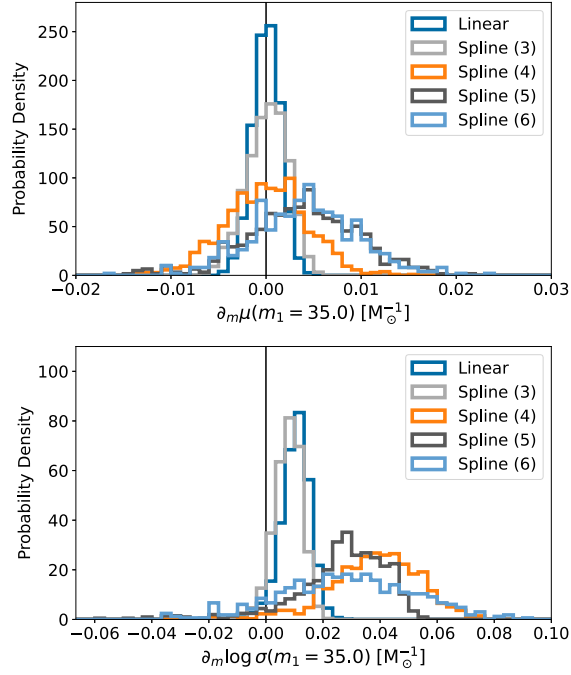
FIG. 8.   Posteriors on the derivative of the mean and width of the Gaussian as a function of primary mass. Each posterior is broadly consistent, with varying levels of confidence that the slope is greater than zero. The derivative of the mean at $m_0 = 35 M_\odot$ has nearly equal support for being positive or negative, while the derivative of the width is positive at $\sim$85%–99.7% confidence for each model.
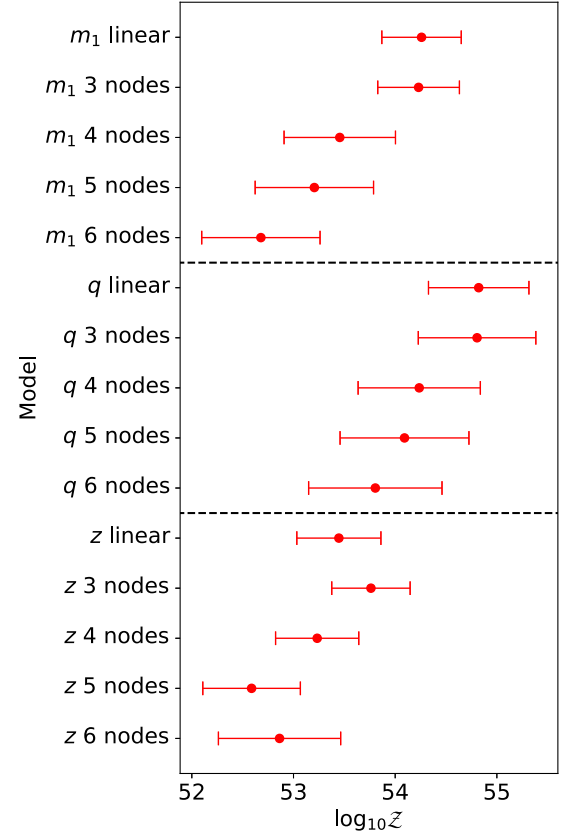

FIG. 9.   Evidences of the GWTC-3 catalog given each of the $\chi_{\rm eff}$ correlation models. Evidence uncertainties include the average Monte Carlo uncertainty intrinsic to the likelihood estimator, added in quadrature with the nested sampling uncertainty of DYNESTY.

TABLE III.   Credibility of a positive slope in the evolution of the mean (denoted $\delta\mu > 0$) at the fiducial value, and of a positive slope in the evolution of the width (denoted $\delta\sigma > 0$). The column on the left is the model assumed, where "Spline (N)" refers to the spline model using N nodes. In each cell, the top line represents the credibility of a positive slope in the mean, and the bottom line represents the same quantity for the width.

	m_1 (at $m_1^* = 35 M_\odot$)	q (at $q^* = 0.9$)	z (at $z^* = 0.2$)
Linear	$p(\partial_m \mu > 0) = 53.9\%$ $p(\partial_m \sigma > 0) = 93.2\%$	$p(\partial_q \mu > 0) = 2.5\%$ $p(\partial_q \sigma > 0) = 41.7\%$	$p(\partial_z \mu > 0) = 16.2\%$ $p(\partial_z \sigma > 0) = 92.7\%$
Spline (3)	$p(\partial_m \mu > 0) = 55.0\%$ $p(\partial_m \sigma > 0) = 89.5\%$	$p(\partial_q \mu > 0) = 13.4\%$ $p(\partial_q \sigma > 0) = 33.0\%$	$p(\partial_z \mu > 0) = 14.2\%$ $p(\partial_z \sigma > 0) = 95.6\%$
Spline (4)	$p(\partial_m \mu > 0) = 60.5\%$ $p(\partial_m \sigma > 0) = 98.7\%$	$p(\partial_q \mu > 0) = 18.7\%$ $p(\partial_q \sigma > 0) = 44.5\%$	$p(\partial_z \mu > 0) = 19.6\%$ $p(\partial_z \sigma > 0) = 98.0\%$
Spline (5)	$p(\partial_m \mu > 0) = 78.3\%$ $p(\partial_m \sigma > 0) = 95.7\%$	$p(\partial_q \mu > 0) = 21.7\%$ $p(\partial_q \sigma > 0) = 58.1\%$	$p(\partial_z \mu > 0) = 17.0\%$ $p(\partial_z \sigma > 0) = 88.8\%$
Spline (6)	$p(\partial_m \mu > 0) = 77.8\%$ $p(\partial_m \sigma > 0) = 93.6\%$	$p(\partial_q \mu > 0) = 30.0\%$ $p(\partial_q \sigma > 0) = 65.5\%$	$p(\partial_z \mu > 0) = 5.9\%$ $p(\partial_z \sigma > 0) = 98.6\%$

APPENDIX C: NONLINEAR CORRELATION IN $\chi_{\rm eff} - z$

To study the potential risk of inferring a nonlinear correlation a GWTC-3-like catalog from a linearly correlated universe, we used 12 sets of 69 detections drawn from an uncorrelated universe observed with Hanford and Livingston in O4-like PSDs [88], in a similar process to the method presented in Sec. IV. These synthetic events were injected from a linearly correlated universe (namely, uncorrelated with slope zero), with a mean $\mu_{\chi_{\rm eff}} = 0.06$ and width $\sigma_{\chi_{\rm eff}} = 0.11$, and the same thresholds, waveforms, and parameter estimation techniques described in Sec. IV. Then, we infer the $\chi_{\rm eff}$ correlation with respect to the redshift z for each synthetic catalog, using the method we introduced in Sec. III. We compute the instantaneous slope at the first three nodes for each hyperposterior sample, and can then estimate the JS divergence between the posteriors of the slopes at neighboring nodes.

We also show the slope posteriors at a fiducial value $z^* = 0.2$ in Fig. 11. We avoid choosing $z^* = 0$, as the data is uninformative in the limit of $z \to 0$ for the same reason it is uninformative for $q \to 0$. While some of the 69 BBH events are closer than others, none of them are consistent with being at $z \to 0$, and indeed it is baked into the population model where the probability density at a given redshift is proportional to the differential comoving volume $p(z) \propto \frac{dV_c}{dz}$ to have zero support at $z \to 0$ [69,114]. In other words, a flexible $z - \chi_{\rm eff}$ correlation model, in particular spline models with appropriate node density and placement, will increase in uncertainty as $z \to 0$. Indeed, we observe this to some degree in each spline model in Fig. 10, but it is especially clear in the spline model with five nodes.
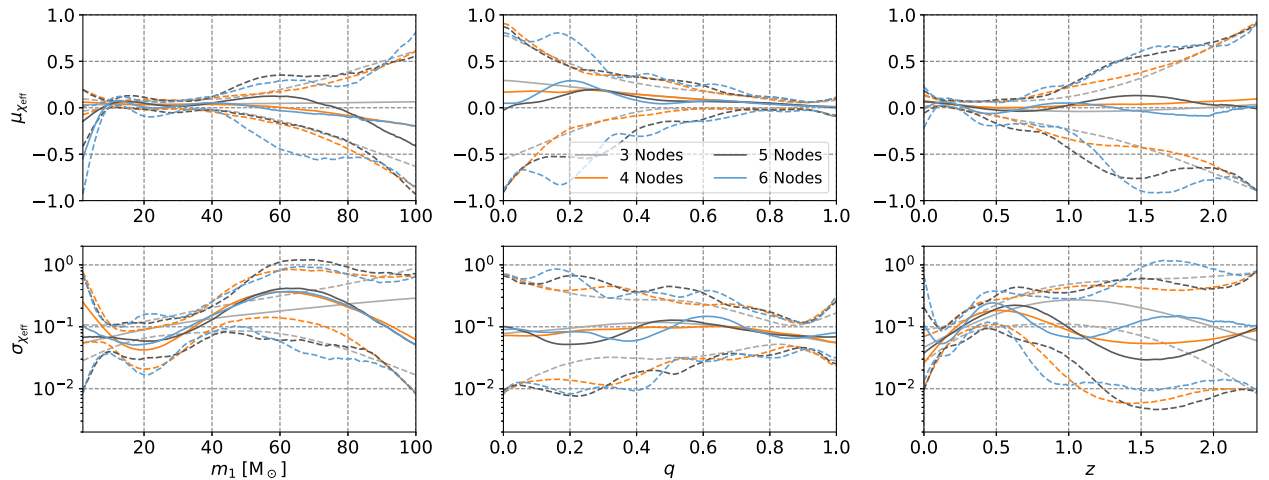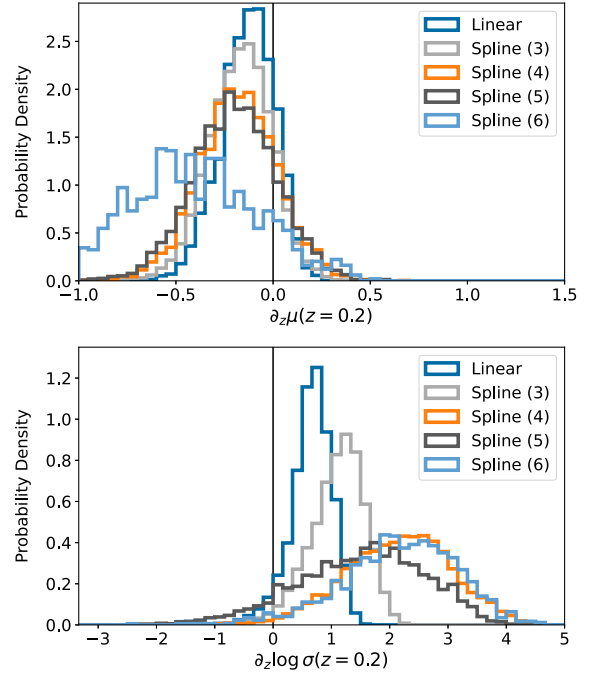

FIG. 11. Posteriors on the derivative of the mean and width of the Gaussian with respect to redshift at $z^* = 0.2$, inferred on GWTC-3 data. Each posterior is consistent, with varying levels of confidence that the slope is greater than zero.

We also should note the significance of broadening with redshift using the linear model is only 92.7%, while Ref. [58] inferred a broadening at 98.6% credibility. We checked various differences between our analyses, and discovered that this difference arises from how selection effects are estimated. In particular, Ref. [58] assumes a semianalytic threshold on the network SNR $\rho_{\rm net} > 9$ for O1 and O2 events, while we use a threshold of $\rho_{\rm net} > 10$, following Ref. [7].

FIG. 10. Spline correlations between $\chi_{\rm eff}$ and each parameter we studied, primary mass m_1 (left column), mass ratio q (center), and redshift z (right). Solid lines represent the median, while the dashed lines represent the boundaries of the central 90% credible interval. The upper panels represent the mean of the $\chi_{\rm eff}$ Gaussian as a function of each parameter, while the lower panels represent the standard deviation of the $\chi_{\rm eff}$ distribution.

APPENDIX D: NODE PLACEMENT AND BIAS IN MOCK CATALOGS

In Sec. IV we study a universe with a nonlinear correlation in the $\chi_{\text{eff}} - q$ plane. We find that, when we fix the x-axis positions of the nodes to be different than the true node positions, we may recover some bias in the inference.

Specifically, the results in Fig. 5 highlight one of the drawbacks for using spline models. Fitting a spline model to a smooth function often results in extraneous structure, bumps that the spline function cannot remove entirely due to its functional form [113]. In the case of the mock catalog we simulated, the correlation is created using a spline node placement (Table II) different from the spline node placement scheme we recover with. Because of this, the spline cannot simultaneously fit the correlation in the region of near equal mass ratio while also effectively fitting the

region of more extreme mass ratios, and so it must compromise with a suboptimal fit across the parameter space. Hence, the spline model "wiggles" when it ideally should be smooth.

Because we are not yet in the limit of infinitely many nodes (we use four nodes in this inference), the spline functions are not perfectly flexible. More nodes are inherently more flexible, and so would presumably do a better job at fitting the correlation. Similarly, using a correct node placement scheme should also improve the fit. Indeed, we find that when we analyze the mock catalog using the same node positions as the true correlation (Table II), the inferred correlation is closer to the truth. This highlights the importance of either running with multiple node counts and placement schemes, or marginalizing over the node x-axis positions as well.

---

[1] R. Abbott *et al.* (LIGO Scientific, VIRGO, and KAGRA Collaborations), 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 *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[3] F. Acernese *et al.* (VIRGO Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[4] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA Collaboration), Interferometer design of the KAGRA gravitational wave detector, Phys. Rev. D **88**, 043007 (2013).

[5] K. Somiya (KAGRA Collaboration), Detector configuration of KAGRA: The Japanese cryogenic gravitational-wave detector, Classical Quantum Gravity **29**, 124007 (2012).

[6] T. Akutsu *et al.* (KAGRA Collaboration), Overview of KAGRA: Detector design and construction history, Prog. Theor. Exp. Phys. **2021**, 05A101 (2021).

[7] R. Abbott *et al.* (KAGRA, VIRGO, and LIGO Scientific Collaborations), Population of merging compact binaries inferred using gravitational waves through GWTC-3, Phys. Rev. X **13**, 011048 (2023).

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

[9] 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).

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

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

[12] M. Fishbach and D. E. Holz, Picky partners: The pairing of component masses in binary black hole mergers, Astrophys. J. Lett. **891**, L27 (2020).

[13] A. M. Farah, M. Fishbach, and D. E. Holz, Two of a kind: Comparing big and small black holes in binaries with gravitational waves, Astrophys. J. **962**, 69 (2024).

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

[15] S. Biscoveanu, M. Isi, S. Vitale, and V. Varma, New spin on LIGO-Virgo binary black holes, Phys. Rev. Lett. **126**, 171103 (2021).

[16] 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).

[17] 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).

[18] 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).

[19] 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).

[20] 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).

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

[22] 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).

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

[24] 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).

[25] M. Zevin, S. S. Bavera, C. P. L. Berry, V. Kalogera, T. Fragos, P. Marchant, C. L. Rodriguez, F. Antonini, D. E. Holz, and C. Pankow, One channel to rule them all? Constraining the origins of binary black holes using multiple formation pathways, Astrophys. J. **910,** 152 (2021).

[26] A. Q. Cheng, M. Zevin, and S. Vitale, What you don't know can hurt you: Use and abuse of astrophysical models in gravitational-wave population analyses, Astrophys. J. **955,** 127 (2023).

[27] K. W. K. Wong, K. Breivik, K. Kremer, and T. Callister, Joint constraints on the field-cluster mixing fraction, common envelope efficiency, and globular cluster radii from a population of binary hole mergers via deep learning, Phys. Rev. D **103,** 083021 (2021).

[28] V. Baibhav, Z. Doctor, and V. Kalogera, Dropping anchor: Understanding the populations of binary black holes with random and aligned-spin orientations, Astrophys. J. **946,** 50 (2023).

[29] J. Godfrey, B. Edelman, and B. Farr, Cosmic cousins: Identification of a subpopulation of binary black holes consistent with isolated binary evolution, arXiv:2304 .01288.

[30] Y.-Z. Wang, Y.-J. Li, J. S. Vink, Y.-Z. Fan, S.-P. Tang, Y. Qin, and D.-M. Wei, Potential subpopulations and assembling tendency of the merging black holes, Astrophys. J. Lett. **941,** L39 (2022).

[31] 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).

[32] 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 (2021).

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

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

[35] D. Foreman-Mackey, D. W. Hogg, and T. D. Morton, Exoplanet population inference and the abundance of earth analogs from noise, incomplete catalogs, Astrophys. J. **795,** 64 (2014).

[36] A. Ray, I. Magaña Hernandez, S. Mohite, J. Creighton, and S. Kapadia, Nonparametric inference of the population of

compact binaries from gravitational-wave observations using binned Gaussian processes, Astrophys. J. **957,** 37 (2023).

[37] T. A. Callister and W. M. Farr, A parameter-free tour of the binary black hole population, arXiv:2302.07289.

[38] I. M. Romero-Shaw, E. Thrane, and P. D. Lasky, When models fail: An introduction to posterior predictive checks and model misspecification in gravitational-wave astronomy, Pub. Astron. Soc. Aust. **39,** e025 (2022).

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

[40] 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).

[41] B. McKernan, K. E. S. Ford, W. Lyra, and H. B. Perets, Intermediate mass black holes in AGN disks: I. Production & Growth, Mon. Not. R. Astron. Soc. **425,** 460 (2012).

[42] N. C. Stone, B. D. Metzger, and Z. Haiman, Assisted inspirals of stellar mass black holes embedded in AGN discs: solving the "final au problem," Mon. Not. R. Astron. Soc. **464,** 946 (2017).

[43] B. McKernan, K. E. S. Ford, R. O'Shaughnessy, and D. Wysocki, Monte Carlo simulations of black hole mergers in AGN discs: Low χ_{eff} mergers and predictions for LIGO, Mon. Not. R. Astron. Soc. **494,** 1203 (2020).

[44] B. McKernan, K. E. S. Ford, T. Callister, W. M. Farr, R. O'Shaughnessy, R. Smith, E. Thrane, and A. Vajpeyi, LIGO–Virgo correlations between mass ratio and effective inspiral spin: Testing the active galactic nuclei channel, Mon. Not. R. Astron. Soc. **514,** 3886 (2022).

[45] H. Tagawa, Z. Haiman, and B. Kocsis, Formation and evolution of compact object binaries in AGN disks, Astrophys. J. **898,** 25 (2020).

[46] S. Mukherjee, The redshift dependence of black hole mass distribution: Is it reliable for standard sirens cosmology?, Mon. Not. R. Astron. Soc. **515,** 5495 (2022).

[47] M. Zevin and S. S. Bavera, Suspicious siblings: The distribution of mass and spin across component black holes in isolated binary evolution, Astrophys. J. **933,** 86 (2022).

[48] C. Karathanasis, S. Mukherjee, and S. Mastrogiovanni, Binary black holes population and cosmology in new lights: Signature of PISN mass and formation channel in GWTC-3, Mon. Not. R. Astron. Soc. **523,** 4539 (2023).

[49] K. Belczynski 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).

[50] S. F. Portegies Zwart and S. L. W. McMillan, The runaway growth of intermediate-mass black holes in dense star clusters, Astrophys. J. **576,** 899 (2002).

[51] C. L. Rodriguez, M. Morscher, B. Pattabiraman, S. Chatterjee, C.-J. Haster, and F. A. Rasio, Binary black hole mergers from globular clusters: Implications for Advanced LIGO, Phys. Rev. Lett. **115,** 051101 (2015); **116,** 029901(E) (2016).

[52] D. Gerosa and M. Fishbach, Hierarchical mergers of stellar-mass black holes and their gravitational-wave signatures, Nat. Astron. **5,** 8 (2021).

[53] 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).

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

[55] C. Adamcewicz, P. D. Lasky, and E. Thrane, Evidence for a correlation between binary black hole mass ratio and black hole spins, Astrophys. J. **958,** 13 (2023).

[56] M. Safarzadeh, W. M. Farr, and E. Ramirez-Ruiz, A trend in the effective spin distribution of LIGO binary black holes with mass, Astrophys. J. **894,** 129 (2020).

[57] G. Franciolini and P. Pani, Searching for mass-spin correlations in the population of gravitational-wave events: The GWTC-3 case study, Phys. Rev. D **105,** 123024 (2022).

[58] 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).

[59] S. Rinaldi, W. Del Pozzo, M. Mapelli, A. L. Medina, and T. Dent, Evidence for the evolution of black hole mass function with redshift, arXiv:2310.03074.

[60] 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).

[61] T. J. Loredo, Accounting for source uncertainties in analyses of astronomical survey data, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by R. Fischer, R. Preuss, and U. V. Toussaint, AIP Conf. Proc. No. 735 (AIP, New York, 2004), p. 195.

[62] W. M. Farr, J. R. Gair, I. Mandel, and C. Cutler, Counting and confusion: Bayesian rate estimation with multiple populations, Phys. Rev. D **91,** 023005 (2015).

[63] 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).

[64] S. Vitale, One, no one, and one hundred thousand—Inferring the properties of a population in presence of selection effects, arXiv:2007.05579.

[65] R. Essick and M. Fishbach, Ensuring consistency between noise and detection in hierarchical Bayesian inference, Astrophys. J. **962,** 169 (2024).

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

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

[68] 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).

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

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

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

[72] P. Ajith *et al.*, Inspiral-merger-ringdown waveforms for black-hole binaries with nonprecessing spins, Phys. Rev. Lett. **106,** 241101 (2011).

[73] S. Vitale, R. Lynch, V. Raymond, R. Sturani, J. Veitch, and P. Graff, Parameter estimation for heavy binary-black holes with networks of second-generation gravitational-wave detectors, Phys. Rev. D **95,** 064053 (2017).

[74] M. Pürrer, M. Hannam, and F. Ohme, Can we measure individual black-hole spins from gravitational-wave observations?, Phys. Rev. D **93,** 084042 (2016).

[75] J. Roulet and M. Zaldarriaga, Constraints on binary black hole populations from LIGO–Virgo detections, Mon. Not. R. Astron. Soc. **484,** 4216 (2019).

[76] C. de Boor, On cubic spline functions that vanish at all knots, Adv. Math. **20,** 1 (1976).

[77] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, Parameter estimation sample release for GWTC-1, Technical Report No. P1800370, LVK, 2020, https://dcc.ligo.org/LIGO-P1800370/public.

[78] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, GWTC-2 data release: Parameter estimation samples and skymaps, Technical Report No. P2000223, LVK, 2021, https://dcc.ligo.org/P2000223/public.

[79] LIGO Scientific Collaboration and Virgo Collaboration, GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run—Parameter estimation data release, 10.5281/zenodo.6513631 (2022).

[80] 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—Parameter estimation data release, 10.5281/zenodo.8177023 (2023).

[81] 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, 10.5281/zenodo.7890398 (2023).

[82] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, Parallelized inference for gravitational-wave astronomy, Phys. Rev. D **100,** 043030 (2019).

[83] J. S. Speagle, dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493,** 3132 (2020).

[84] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. Ser. **241,** 27 (2019).

[85] This is somewhat at odds with the maximum redshift considered in the sensitivity injections of [81], with

$z_{max} = 1.9$. We verified our results are unchanged after considering this adjustment.

[86] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22,** 79 (1951).

[87] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory **37,** 145 (1991).

[88] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, Noise curves used for Simulations in the update of the Observing Scenarios Paper, Technical Report No. T2000012, LVK, 2022, https://dcc.ligo.org/LIGO-T2000012/public.

[89] G. Pratten et al., Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103,** 104056 (2021).

[90] N. J. Cornish, Fast fisher matrices and lazy likelihoods, arXiv:1007.4820.

[91] N. J. Cornish, Heterodyned likelihood for rapid gravitational wave parameter inference, Phys. Rev. D **104,** 104054 (2021).

[92] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, arXiv:1806.08792.

[93] R. Essick, Semianalytic sensitivity estimates for catalogs of gravitational-wave transients, Phys. Rev. D **108,** 043011 (2023).

[94] There is an issue here with nomenclature. Often, the matched-filter SNR is defined as a *maximization* over a fixed signal bank, however for synthetic catalogs it is typical to compute the matched-filter SNR with the signal from the true source parameters. These statistics are conditioned on different things: the former is conditioned only on the data, because the signal bank is fixed, while the latter is conditioned on the data *and* the true signal parameters.

[95] F. S. Broekgaarden, S. Stevenson, and E. Thrane, Signatures of mass ratio reversal in gravitational waves from merging binary black holes, Astrophys. J. **938,** 45 (2022).

[96] S. S. Bavera et al., The impact of mass-transfer physics on the observable properties of field binary black hole populations, Astron. Astrophys. **647,** A153 (2021).

[97] A. Santini, D. Gerosa, R. Cotesta, and E. Berti, Black-hole mergers in disklike environments could explain the observed q-χeff correlation, Phys. Rev. D **108,** 083033 (2023).

[98] 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).

[99] S. S. Bavera, M. Zevin, and T. Fragos, Approximations to the spin of close Black-hole-Wolf-Rayet binaries, arXiv:2105.09077.

[100] J. Fuller and W. Lu, The spins of compact objects born from helium stars in binary systems, Mon. Not. R. Astron. Soc. **511,** 3951 (2022).

[101] 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).

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

[103] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, Illuminating black hole binary formation channels with spins in Advanced LIGO, Astrophys. J. Lett. **832,** L2 (2016).

[104] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, Spin orientations of merging black holes formed from the evolution of stellar binaries, Phys. Rev. D **98,** 084036 (2018).

[105] T. A. Callister, W. M. Farr, and M. Renzo, State of the field: Binary black hole natal kicks and prospects for isolated field formation after GWTC-2, Astrophys. J. **920,** 157 (2021).

[106] S. Stevenson, Biases in estimates of black hole kicks from the spin distribution of binary black holes, Astrophys. J. Lett. **926,** L32 (2022).

[107] D. Gerosa and E. Berti, Are merging black holes born from stellar collapse or previous mergers?, Phys. Rev. D **95,** 124046 (2017).

[108] M. Fishbach, D. E. Holz, and B. Farr, Are LIGO's black holes made from smaller black holes?, Astrophys. J. Lett. **840,** L24 (2017).

[109] H. Tagawa, Z. Haiman, I. Bartos, B. Kocsis, and K. Omukai, Signatures of hierarchical mergers in black hole spin and mass distribution, Mon. Not. R. Astron. Soc. **507,** 3362 (2021).

[110] M. Fishbach, C. Kimball, and V. Kalogera, Limits on hierarchical black hole mergers from the most negative χ_{eff} systems, Astrophys. J. Lett. **935,** L26 (2022).

[111] Z. Doctor, D. Wysocki, R. O'Shaughnessy, D. E. Holz, and B. Farr, Black hole coagulation: Modeling hierarchical mergers in black hole populations, arXiv:1911.04424.

[112] C. Kimball, C. Talbot, C. P. L. Berry, M. Carney, M. Zevin, E. Thrane, and V. Kalogera, Black hole genealogy: Identifying hierarchical mergers with gravitational waves, Astrophys. J. **900,** 177 (2020).

[113] A. M. Farah, B. Edelman, M. Zevin, M. Fishbach, J. M. 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).

[114] D. W. Hogg, Distance measures in cosmology, arXiv:astro-ph/9905116.

Correction: Errors in the author list and their affiliations have been set right and a missing footnote for the second author has been added.