arXiv:1910.10175v1 [astro-ph.GA] 22 Oct 2019

Constraining black hole-galaxy scaling relations from the large-scale clustering of Active Galactic Nuclei and implied mean radiative efficiency

Francesco Shankar[1], Viola Allevato[2,3,4], Mariangela Bernardi[5], Christopher Marsden[1], Andrea Lapi[6,7], Nicola Menci[8], Philip J. Grylls[1], Mirko Krumpe[9], Lorenzo Zanisi[1], Federica Ricci[10], Fabio La Franca[11], Ranieri D. Baldi[1], Jorge Moreno[12,13], Ravi K. Sheth[5]

[1]*Department of Physics and Astronomy, University of Southampton, Highfield, SO17 1BJ, UK*
[2]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[3]*Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2a, 00014 University of Helsinki, Finland*
[4]*INAF-Osservatorio Astronomico di Bologna, Via Ranzani 1, 40127 Bologna, Italy*
[5]*Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd St, Philadelphia, PA 19104, USA*
[6]*SISSA, Via Bonomea 265, I-34136 Trieste, Italy*
[7]*IFPU - Institute for fundamental physics of the Universe, Via Beirut 2, 34014 Trieste, Italy*
[8]*INAF–Osservatorio Astronomico di Roma, via Frascati 33, 00078 Monteporzio Catone, Italy*
[9]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[10]*Instituto de Astrofísica and Centro de Astroingeniería, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile*
[11]*Dipartimento di Matematica e Fisica, Università Roma Tre, via della Vasca Navale 84, I-00146 Roma, Italy*
[12]*Department of Physics and Astronomy, Pomona College, Claremont, CA 91711, USA*
[13]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA*

A supermassive black hole has been found at the centre of nearly every galaxy observed with sufficient sensitivity. The masses of these black holes are observed to increase with either the total mass or the mean (random) velocity of the stars in their host galaxies. The origin of these correlations remains elusive. Observational systematics and biases severely limit our knowledge of the local demography of supermassive black holes thus preventing accurate model comparisons and progress in this field. Here we show that the large-scale spatial distribution of local active galactic nuclei (AGN), believed to be accreting supermassive black holes, can constrain the shape and normalization of the black hole-stellar mass relation thus bypassing resolution-related observational biases. In turn, our results can set more stringent constraints on the so-called "radiative efficiency", ε, a fundamental parameter describing the inner physics of supermassive black holes that is closely linked to their spin, geometry, and ability to release energy. The mean value of ε can be estimated by comparing the average total luminous output of AGN with the relic mass density locked up in quiescent supermassive black holes at galaxy centres today. For currently accepted values of the AGN obscured

fractions and bolometric corrections, our newest estimates of the local supermassive black hole mass density favour mean radiative efficiencies of $\varepsilon \sim 10 - 20\%$, suggesting that the vast majority of supermassive black holes are spinning moderately to rapidly. With large-scale AGN surveys coming online, our novel methodology will enable even tighter constraints on the fundamental parameters that regulate the growth of supermassive black holes.

In the traditional picture, quiescent supermassive black holes at the centres of local galaxies today are the relics of a previous phase of active gas accretion in which they shone as Active Galactic Nuclei (AGN)[1]. The brightness of an AGN depends on how efficiently the gas orbiting the black hole can radiate. In principle, the mean radiative efficiency ε (and hence black hole spin[2]) can be constrained by equating[3] the mass accreted by all AGN over all times with the mass density in supermassive black holes today[4–7]. The accreted mass scales roughly with the inverse of the mean radiative efficiency, because a lower radiative efficiency requires a higher mass accretion rate to produce the observed AGN luminosity. In this scenario, the same total AGN emissivity can be matched either by a small black hole mass density today but a large radiative efficiency, or vice-versa. This has fueled interest in estimating the local black hole mass density or "mass function", i.e. the number density per unit comoving volume and bin of black hole mass.

Quiescent black holes are difficult to observe, so large samples are not available. Therefore, the black hole mass function is usually estimated in two steps. First, one uses the small sample that is available to calibrate how a black hole's mass correlates with other properties of its host galaxy, such as its stellar mass $M_{\rm star}$ or the velocity dispersion of its stars σ. Since $M_{\rm star}$ and σ can be estimated in much larger samples, one uses these "scaling relations" to transform the comoving number density of $M_{\rm star}$ or σ into that for black hole mass[4].

Unfortunately, despite its elegance, the method described above suffers from systematic uncertainties that prevent a robust estimate of ε, the mean radiative efficiency of supermassive black holes. The main difficulty is to estimate the local scaling relations reliably[8]. Over the past decade, a number of groups have put forward a variety of rather different relations. For instance, there are several claims that black holes hosted in galaxies belonging to different morphological classes follow different scaling relations[9,10]. In addition, active and quiescent black holes in the local Universe appear to follow somewhat differently normalized scaling relations, with the degree of the offset depending on the type of galaxies, AGN sample, or scaling relation considered[11–13]. Furthermore, limitations on instrumental capabilities or other observational effects may systematically bias the observed scaling relations away from the intrinsic ones. This is a concern especially in quiescent early-type galaxies with dynamically-measured black hole masses[8,14,15]. In fact, recent work[13] suggests that such a bias may explain much of the reported difference between active and quiescent scaling relations: the intrinsic relations may be much more similar than they appear.

In what follows, we present a methodology that uses AGN clustering to set new and valuable *independent* constraints on the overall shape of the local black hole scaling relations. As we describe below, *simultaneously* matching the large-scale AGN clustering and the local black hole

mass density provides more robust constraints on the mean black hole radiative efficiency ε.

The connection between AGN clustering, scaling relations and radiative efficiency

Within the framework of a cold dark matter Universe, more massive host dark matter haloes appear progressively more strongly clustered, i.e., their spatial distribution shows more marked departures from an underlying random distribution[16]. Galaxies and black holes residing in more massive dark matter haloes are thus naturally expected to appear more clustered. Our methodology builds on this basic notion of galaxy clustering and it can be briefly outlined as follows:

- At any redshift of interest z, we first create large catalogues of host dark matter haloes from the halo mass function $n(M_{\rm halo})$. In practice, in a given (large) volume V we select the halo masses of mass $M_{\rm halo}$ for which the cumulative halo mass function is an integer, i.e. $V \times n(> M_{\rm halo}) = N$, with $N \in \mathbb{I}$. To each dark matter halo we then assign a central galaxy with stellar mass given via the $M_{\rm star}[z] - M_{\rm halo}[z]$ relation. The latter relation is inferred from "abundance matching" arguments, i.e., based on number density equivalence between galaxy and host halo number counts[17,18], $n(> M_{\rm star}, z) = n(> M_{\rm halo}, z)$. We include a scatter of 0.15 dex in stellar mass at fixed host halo mass[17].

- To each "mock" galaxy we then assign a central supermassive black hole with mass as given by an input $M_{\rm bh}[z] - M_{\rm star}[z]$ empirical relation. We include a scatter of $\sim 0.4 - 0.5$ dex in black hole mass at fixed stellar mass[8,13], as detailed below. We will also explore a model in which we bypass the $M_{\rm bh}[z] - M_{\rm star}[z]$ relation assigning black holes via the $M_{\rm bh} - \sigma$ relation (with a scatter of 0.3 dex in black hole mass), as explained below.

- The combination between black-hole and galaxy property (stellar mass or velocity dispersion), and galaxy property with host halo mass effectively predicts a $M_{\rm bh}[z] - M_{\rm halo}[z]$ relation.

- More massive haloes are more strongly clustered, and their (large-scale) clustering strength is encoded in a parameter "b" called the bias, as explained below. More massive haloes are on average characterized by larger b-values.

- For a given input $M_{\rm bh}[z] - M_{\rm star}[z]$ relation, we compute the implied $M_{\rm bh}[z] - M_{\rm halo}[z]$ relation, and finally the predicted large-scale clustering of black holes, encoded in the $b - M_{\rm bh}$ relation.

- The higher the normalization in the $M_{\rm bh}[z] - M_{\rm halo}[z]$ relation, the lower is the $M_{\rm halo}$ that hosts a given $M_{\rm bh}$, and so the weaker is the expected clustering strength of black holes at fixed black hole mass (and dispersion around the mean).

- The $M_{\rm bh}[z] - M_{\rm halo}[z]$ relation(s) that provide the closest match to the AGN clustering measurements, will set key constraints on, most noticeably, the most appropriate input $M_{\rm bh}[z] - M_{\rm star}[z]$ relation.

- After assigning to each galaxy a supermassive black hole mass via the $M_{\rm bh}[z] - M_{\rm star}[z]$ relation favoured by AGN clustering, we then estimate the implied $V_{\rm max}$-weighted black hole mass density and mass function.
- Finally, the match between the local mass density and accreted mass density determined from integrated AGN number counts, will yield more secure constraints on the mean radiative efficiency of supermassive black holes.

Results

The local $M_{\rm bh}$-$M_{\rm star}$ relation and implied $M_{\rm bh}$-$M_{\rm halo}$ relation We start by discussing the $M_{\rm bh}$-$M_{\rm star}$ relation between black hole mass and host galaxy (total) stellar mass, which is perhaps the most actively debated correlation between black holes and galaxies (Figure 1) yet widely adopted as a reference in black hole-galaxy co-evolution cosmological models[19]. The left panel of Figure 1 reports the latest renditions of the $M_{\rm bh}$-$M_{\rm star}$ relation (full details on the different scaling relations can be found in the "Methods' Section) from local galaxy samples with dynamically-measured black hole masses from Savorgnan et al.[20] (for their subsample of secure black-hole mass measurements[8,13]), Kormendy & Ho[9], and Sahu et al.[21], as labelled in the top left corner. The same panel also includes, for completeness, the $M_{\rm bh}$-$M_{\rm star}$ relation (blue, triple dot-dashed line) recently derived by Davis et al.[10] for a local sample of 40 spiral galaxies, which appears to be around three times steeper than the one characterizing the observed sample of early-type galaxies (dotted line). Also shown is the $M_{\rm bh}$-$M_{\rm star}$ relation from Reines & Volonteri[12] (dashed, green line) and Baron & Ménard[22] (dot-dashed purple line) extracted, respectively, from a local AGN samples of 244 and 2000 sources with single-epoch black hole mass estimates. The "intrinsic" $M_{\rm bh}$-$M_{\rm star}$ relation proposed by Shankar et al.[8] is reported by a solid red line with its scatter marked by the yellow region. Hereafter, we will refer to the de-biased $M_{\rm bh}$-$M_{\rm star}$ relation put forward by Shankar et al.[8] (solid red line in the left panel of Figure 1, with a scatter of $\lesssim 0.4$ dex, marked by the yellow band) as the "intrinsic" or "unbiased" scaling relation.

The very first feature to highlight in the left panel of Figure 1 is the systematic discrepancy between the fits to the local quiescent samples of supermassive black holes and Shankar et al.'s intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation. Shankar et al.[8] devised targeted Monte Carlo simulations to show that local quiescent (mainly early-type) galaxies having dynamically measured black hole mass, present larger velocity dispersions at fixed stellar mass with respect to galaxies in the Sloan Digital Sky Survey (SDSS). In fact, given the strong and fundamental[8,13,14,28] dependence $M_{\rm bh} \propto \sigma^{4-5}$, higher velocity dispersions would on average select higher mass black holes at fixed host galaxy stellar mass. Thus, the local quiescent samples of (mainly early-type) galaxies with dynamically measured black holes would then follow *biased* scaling relations, with fictitiously higher normalizations than the bulk of the underlying population of supermassive black holes. To a large extent, Shankar et al.[8] showed that a possible reason behind the higher velocity dispersions in the local sample of quiescent early-type galaxies could be ascribed to a bias arising from the requirement that the central black hole's sphere of influence must be resolved to measure black hole masses

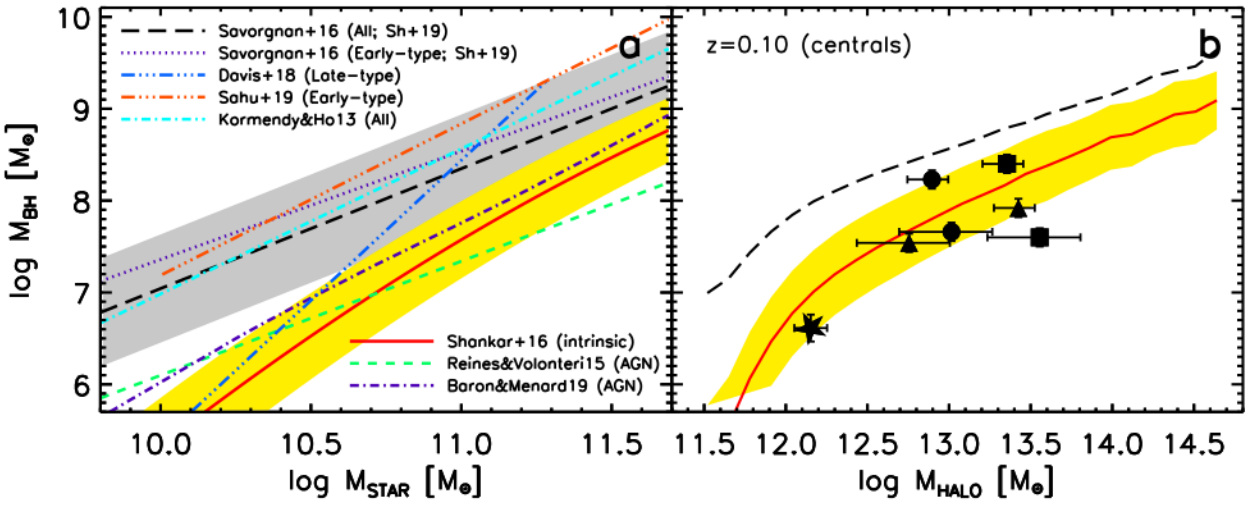

Figure 1: **Overview of local scaling relations between black hole mass and host (total) stellar mass.** *Left*: Correlations between central black hole mass and host galaxy total stellar mass in the local Universe. The black long-dashed line with its scatter (grey band) is the relation inferred from the (whole) quiescent sample by Savorgnan et al.[20], while the purple, dotted line refers to only their early-type galaxies (ellipticals and lenticulars). The triple dot-dashed orange line is the fits to local quiescent samples of early-type galaxies with dynamical measures of black holes by Sahu et al.[21]. The dot-dashed, cyan line is a linear fit to the sample of Kormendy & Ho[9]. The solid red line with its scatter (yellow region) is the unbiased $M_{\rm bh}$-$M_{\rm star}$ relation from Shankar et al.[8]. The green dashed and purple dot-dashed lines are the fits to the local active samples from, respectively, Reines & Volonteri[12] and Baron & Ménard[22], while the blue triple-dot line is the fit to a local sample of late-type galaxies by Davis et al.[10] with dynamically-measured black hole masses. *Right*: Correlations between black hole mass host halo mass at $z = 0.1$ (halos are defined to be 200 times the background density) implied by the $M_{\rm star}$-$M_{\rm halo}$ relation extracted from abundance matching coupled to the $M_{\rm bh}$-$M_{\rm star}$ observed and unbiased relations (black long-dashed and red solid lines in the left panel). Data are from Powell et al.[23] (squares), Krumpe et al.[24,25](triangles and circles, respectively) and the Milky Way black hole[26,27] (star).

with spatially resolved kinematics.

As already noted by Reines & Volonteri, the relations defined by the AGN samples are about an order of magnitude lower than the ones defined by the quiescent early-type samples, in line with several other previous claims[11,29,30]. In fact, Shankar et al. (2019)[13] have highlighted the fact that the vast majority of local AGN samples follow relations that are noticeably lower in normalization than those of quiescent galaxies and closer to the intrinsic relation (yellow region). They argue that AGN do not suffer from angular resolution-selection effects, so they may more faithfully trace the intrinsic black hole-galaxy scaling relations.

In summary, different host galaxy morphology, observational biases and selection effects all tend to result in different black hole mass–host galaxy scaling relations. To make progress, we need an independent constraint on these scaling relations, which we obtain as follows.

It is well known that galaxies with large stellar masses are more strongly clustered[16]. Therefore, it should be possible to set constraints on the normalization and steepness of the $M_{\rm bh}$-$M_{\rm star}$ relation by including *independent* galaxy clustering measurements. The right panel of Figure 1 reports the $z = 0.1$ implied $M_{\rm bh}$-$M_{\rm halo}$ relation of central galaxies obtained by assigning to each dark matter halo extracted from the parent halo mass function[31], a galaxy stellar mass from the mean $M_{\rm star}$-$M_{\rm halo}$ relation (and its scatter), and then a black hole mass from the $M_{\rm bh}$-$M_{\rm star}$ relation (with scatter) for the quiescent Savorgnan et al.[20] sample (black long-dashed line) and the intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation from Shankar et al. (2016; solid red line and its scatter marked by the yellow region)[8]. The latter $M_{\rm bh}$-$M_{\rm star}$ relation clearly predicts a correlation between black hole mass and host halo mass that is more consistent with the variety of data extracted from low-redshift AGN large-scale optical surveys[23–25] and reported in the right panel of Figure 1. In what follows we always adopt as a reference for the "biased/observed" $M_{\rm bh}$-$M_{\rm star}$ relation the one inferred from the Savorgnan et al. sample. As inferred from the left panel of Figure 1, the Savorgnan et al. relation is the shallowest and the lowest in normalization among the various renditions of the dynamically-based $M_{\rm bh}$-$M_{\rm star}$ scaling relations. Adopting any of the other $M_{\rm bh}$-$M_{\rm star}$ relation characterizing quiescent black hole samples[21] would, if anything, further exacerbate the discrepancy with AGN clustering measurements.

Independent constraints on the $M_{\rm bh}$-$M_{\rm star}$ relation from AGN large-scale clustering To further improve on the constraints that can be derived from AGN samples, we now use the large-scale AGN clustering measurements performed on the ROSAT All-Sky Survey and SDSS data sets[24] in the redshift range $0.16 < z < 0.36$. AGN clustering is usually quantified by the two-point correlation function: the excess number of AGN pairs in the data compared to a random distribution of the same number density, as a function of separation. At scales larger than a few Mpc, we are in the linear regime of structure formation and the ratio of the two-point correlation function to that expected for the dark matter is approximately independent of pair separation. The square root of this ratio is called the "large-scale bias factor" b, and it is not to be confused with the bias in the

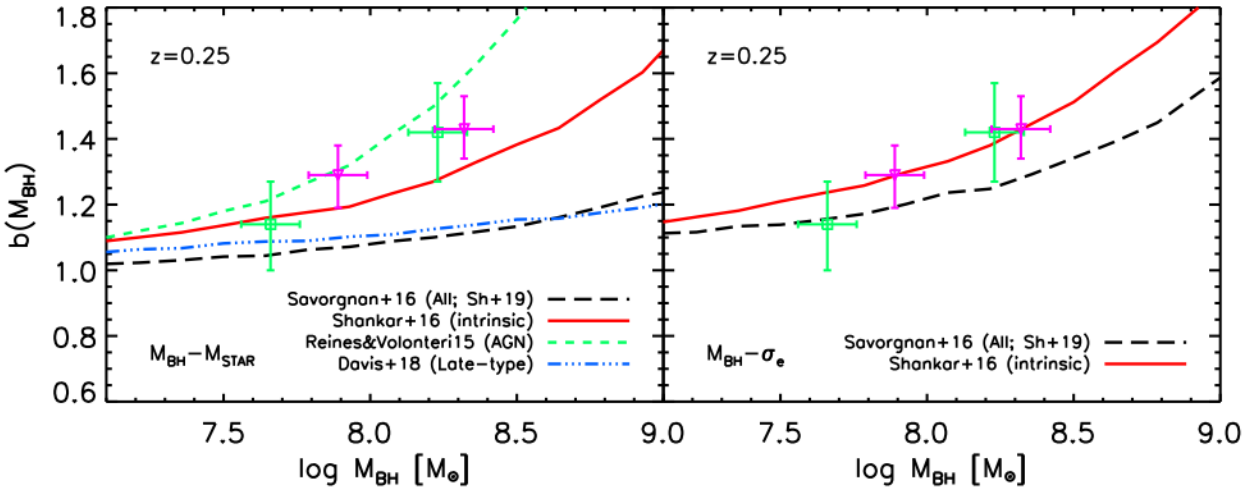

Figure 2: **Predicted bias as a function of black hole mass.** Left: Results for the mean large-scale clustering are shown at $z = 0.25$ for unbiased and observed (red solid and long-dashed black lines, respectively), the Reines & Volonteri[12] (dashed green line), and the (late-type) Davis et al.[10] (blue triple-dot-dashed line) $M_{\rm bh}$-$M_{\rm star}$ relations, as labelled. The data[24] (green squares and purple triangles are X-ray and optical AGN, respectively) are extracted from the clustering properties of AGN identified in ROSAT and SDSS in the redshift range $0.16 < z < 0.36$. The models with the unbiased/lower normalization $M_{\rm bh}$-$M_{\rm star}$ relation are favoured by current data. Right: Similar to left hand panel, but now showing $b(M_{\rm bh})$ expected from the observed/biased (dashed) and intrinsic (solid) $M_{\rm bh}$-σ scaling relations.

scaling relations we were discussing about in the previous Sections! The large-scale bias factor b differs for different galaxy samples. The bias factor in fact progressively increases for larger halo masses[16]. Due to the monotonic dependence between stellar mass and host halo mass in the $M_{\rm star}$-$M_{\rm halo}$ relation, especially at the masses of interest in this work (intrinsic scatter of 0.15 dex), samples with larger $M_{\rm star}$, will on average appear more strongly clustered and characterized by a larger bias factor b[16]. Since black hole mass is also monotonically related to host halo mass (albeit with a larger scatter than in the $M_{\rm star}$-$M_{\rm halo}$ relation), we would thus expect the b factor to increase with black hole mass.

The two green squares with error bars in Figure 2 (same in both panels) show the large-scale bias factors for two samples of low- and high-mass active supermassive black holes observed in X-ray band. Similarly, the two magenta triangles show the bias factor b in a low- and high-mass sample of optically-selected AGN. The inferred bias factors b appear to be rather independent of wavelength, and indeed larger for the samples with larger black hole mass $M_{\rm bh}$.

The curves in the left-hand panel of Figure 2 show the expected large-scale bias assuming black hole masses are assigned via the observed/biased (long-dashed black, Equation 1) or intrinsic/unbiased (solid red) $M_{\rm bh}$-$M_{\rm star}$ scaling relations[8] (Equations 2 and 3), with the same underlying mean $M_{\rm star}$-$M_{\rm halo}$ relation (see the "Methods" Section). The predicted bias parameter b based on the observed/biased scaling relations is too low and is ruled out at high confidence ($\gtrsim 3\sigma$), whilst that based on the intrinsic/unbiased scaling relations provides a good match to the data. The agreement is remarkable, given that there are no free parameters to fit. In the Methods Section we provide a detailed comparison with the measured cross-correlation of AGN-SDSS sources and careful estimates of the significance of the discrepancies between models and data. The dashed green curve shows that the Reines & Volonteri[12] $M_{\rm bh}$-$M_{\rm star}$ scaling relation also yields a good match to the clustering measurements, implying that the AGN clustering measurements are still consistent with $M_{\rm bh}$-$M_{\rm star}$ relations that are even lower in normalization than Shankar et al.'s intrinsic relation, further strengthening the key results of this work. For the Reines & Volonteri relation we assume a nominal intrinsic scatter smaller of 0.3 dex, of the order of what inferred by the Authors, but slightly smaller than the ones characterizing the scaling relations of dynamically-measured inactive black holes[13].

It has been argued that host galaxy morphology may play a substantial role in creating the systematic offset between black hole scaling relations of active and quiescent galaxies. In fact the former active black holes could be mostly hosted in later-type galaxies, which are expected to allegedly be hosting lower black hole masses at fixed stellar mass, while the latter quiescent samples with larger black hole masses tend to inhabit earlier-type galaxies[12]. However, the Davis et al.[10] steep relation of dynamically-measured quiescent black holes in late-type galaxies (blue triple dot-dashed line, with an intrinsic scatter of 0.7 dex), tends to still fall substantially below the bias data at high $M_{\rm bh}$. The latter result clearly supports the view of a bias in the observed $M_{\rm bh}$-$M_{\rm star}$ relation which is independent of galaxy morphology or AGN type. We also expect the Krumpe et al. AGN clustering measurements, based on large, serendipitous samples, to be

representative of a variety of host galaxy morphologies and not just late types. We finally note that all of the data in Figure 2 mostly refer to Type 1 AGN. Observationally, Type 2/obscured AGN have anyhow been measured to cluster at a comparable or even higher level, at least at $z < 1$[32,33], which would further strengthen our results in favour of lower normalizations in the $M_{\rm bh}$-$M_{\rm star}$ intrinsic relation.

In summary, the low-redshift AGN clustering measurements considered here favour considerably *lower* normalizations of the $M_{\rm bh}$-$M_{\rm star}$ relation. Scalings in $M_{\rm bh}$-$M_{\rm star}$ with high normalizations such as the observed/biased one characterizing quiescent galaxies (black long-dashed line), tend to map black holes to lower stellar masses and thus to lower host halo masses (see Figure 1), significantly decreasing the predicted clustering signal. On the other hand, $M_{\rm bh}$-$M_{\rm star}$ scalings such as those from the intrinsic relation (red solid line) or from the AGN sample (green dashed line), better line up with the clustering data. AGN clustering thus provides additional, independent evidence for the presence of a bias in the observed $M_{\rm bh}$-$M_{\rm star}$ relation, at least for quiescent black holes with $M_{\rm bh} \gtrsim 3 \times 10^7\,M_{\odot}$. At lower black hole masses AGN clustering data that are currently available lose their constraining power. For example, for $M_{\rm bh} \lesssim 3 \times 10^7\,M_{\odot}$, AGN clustering data are not able to distinguish between the steep (late-type) relation of Davis et al. or the flatter one of Reines & Volonteri. This loss in constraining power is a consequence of the fact that $b(M_{\rm halo})$ becomes nearly constant at low masses[35].

For completeness, the right-hand panel of Figure 2 shows the bias parameter $b(M_{\rm bh})$ expected from the $M_{\rm bh}$-σ scaling relation. Measuring the clustering strength implied by the $M_{\rm bh}$-σ relation represents an additional independent test to the overall reliability of the black hole scaling relations. This test is also particularly valuable as it circumvents the use of $M_{\rm star}$ by directly applying abundance matching between the velocity dispersion-based local black hole mass function and halo (plus subhalo) mass functions (see the "Methods" Section for full details). Moreover, it has been claimed that the $M_{\rm bh}$-σ relation is less biased than the $M_{\rm bh}$-$M_{\rm star}$ relation, implying that the former should be closer to the "intrinsic" relation, in terms of normalization, slope, dispersion around the mean[8,13,14]. It is thus expected that the related (large-scale) clustering properties should be less sensitive to observational biases. Indeed, when adopting reasonable scatter (~ 0.3 dex) around the mean $M_{\rm bh}$-$M_{\rm halo}$ relation[7,34], when adopting the observed (dashed black) and intrinsic (solid red) $M_{\rm bh}$-σ relations, in both cases we find mean b-$M_{\rm bh}$ relations broadly consistent with the data, with the former only slightly disfavoured at the highest black hole masses.

Comparing local and accreted mass functions: Constraining the mean radiative efficiency of supermassive black holes We first compute the local mass function of supermassive black holes from the SDSS galaxy sample adopted by Shankar et al.[8] (full details given in the Methods Section). The 1σ error in number densities at fixed black hole mass are shown in Figure 3 for the unbiased and biased scaling relations, respectively. The left panel of Figure 3 points to a large difference between black hole mass functions computed from the observed and intrinsic $M_{\rm bh}$-$M_{\rm star}$ scaling relations (green and cyan areas, respectively; the vertical grey band marks the region in which black hole mass function estimates become progressively less secure). Of uttermost impor-

tance, the black hole mass function derived from the observed $M_{\rm bh}$-σ relation (dashed red line) still presents noticeable departures from the one derived from the observed $M_{\rm bh}$-$M_{\rm star}$ relation (green region). As proven in Figure 5, this behaviour is *not* a consequence of neglecting bulge corrections but it reflects an internal inconsistency in the samples of local quiescent black holes[8,13,14]. On the other hand, the black hole mass function derived from the intrinsic $M_{\rm bh}$-σ relation[8] (long-dashed blue line in Figure 3) is fully consistent with the black hole mass function derived from the intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation (cyan region; Equation 2).

In the right panel of Figure 3 we then compare the local black hole mass functions with the accreted mass functions derived from AGN number counts and continuity equation techniques. We show results for the accreted black hole mass function computed in two ways, via the method outlined in Shankar et al.[6] (long-dashed red lines) and as proposed by Aversa et al.[7] (triple dot-dashed purple lines). For both models we adopt the same input X-ray AGN luminosity function[36] with bolometric correction from Yang et al.[37], and a mean radiative efficiency of $\varepsilon = 0.15$. In all accretion models we include obscured AGN with column densities up to $\log N_H/{\rm cm}^{-2} \lesssim 25$, as observed in a number of studies[38]. The two sets of lines for each model bracket the systematic uncertainty in including or excluding obscured AGN with higher column densities, $25 < \log N_H/{\rm cm}^{-2)} < 26$, following the N_H column density distributions provided by Ueda et al.[36]. Despite the substantially different input Eddington ratio distributions, both accretion models predict similar accreted black hole mass functions that line up well with the local black hole mass function derived from the intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation (cyan area). Broadly matching the local black hole mass function derived from the observed $M_{\rm bh}$-$M_{\rm star}$ relation would instead require at least one order of magnitude lower radiative efficiencies (dotted red lines and yellow area). It is interesting to note the good match between local and accreted mass function even at the largest mass scales. As discussed in previous works[6,7], frequent black hole mergers tend to generate a relatively long high-mass tail in the accreted black hole mass function. The latter would be in tension with the nearly-exponential fall off of the local black hole mass function, unless an appropriately fine-tuned radiative efficiency progressively increasing with black hole mass is adopted[6].

In integral form, Figure 4 shows the comparison between the local black hole and accreted mass *densities*. The former is obtained from direct integration of the local black hole mass function derived from the $M_{\rm bh}$-$M_{\rm star}$ relation (long-dashed black box, corresponding to integration of the cyan region in Figure 3; also included, for completeness, the mean local black hole mass density implied by the intrinsic $M_{\rm bh}$-σ relation, pink square). The latter accreted mass density is retrieved from integration of the AGN bolometric luminosity function over luminosity and cosmic time (see Methods Section). The three coloured bands in each panel of Figure 4 mark three different values of the radiative efficiency, as labelled in each panel. Each coloured band brackets the systematic uncertainty induced by including or neglecting the Compton-thick AGN with $25 < \log N_H/{\rm cm}^{-2} < 26$. The top row includes model outputs that make use of the Yang et al.[37] and Marconi et al.[5] bolometric corrections, respectively. The right panels also allow for an additional loss of rest-mass energy via, e.g., jets and/or winds, parameterized by the kinetic efficiency which we set[39] to an average value of $\varepsilon_{\rm kin} = 0.15$. It is interesting to note that Yang et al.[37] also computed

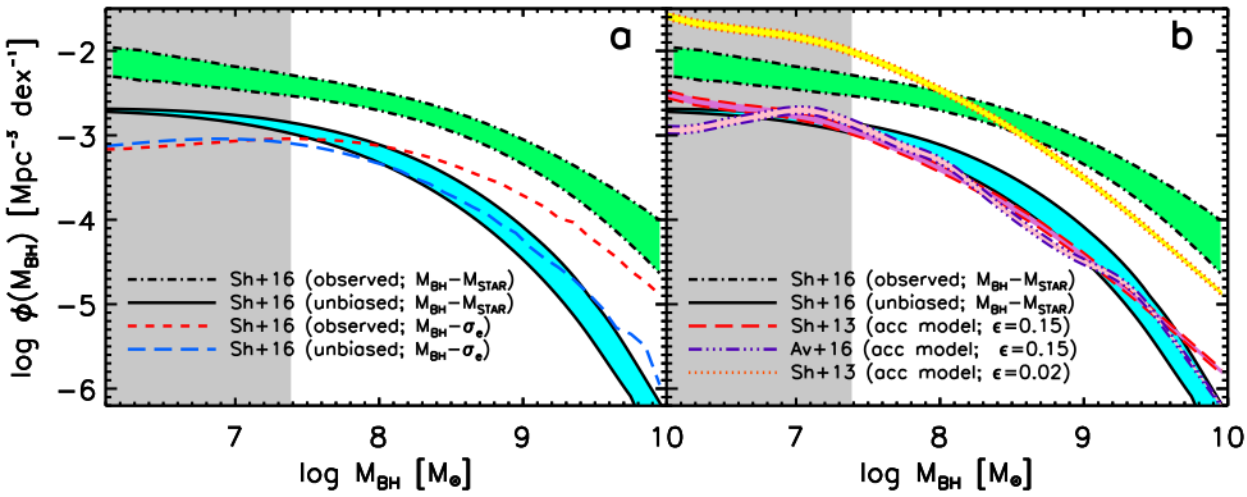

Figure 3: **Comparing local and accreted mass functions.** *Left*: Estimates of the local black hole mass functions as derived from the observed $M_{\rm bh}$-$M_{\rm star}$ relation (dot-dashed lines and green region), the unbiased[8] $M_{\rm bh}$-$M_{\rm star}$ relation (solid black lines and cyan region), the observed $M_{\rm bh}$-σ relation (red dashed line), and the unbiased[8] $M_{\rm bh}$-σ relation (blue long-dashed line). A mismatch between the observed $M_{\rm bh}$-$M_{\rm star}$- and $M_{\rm bh}$-σ-based black hole mass functions (green region and red dashed line, respectively) is evident. *Right*: Comparison between the local black hole mass functions from the left panel with the accreted black hole mass functions at $z = 0.1$ from the Shankar et al.[6] and Aversa et al.[7] accretion models (dotted red and purple triple dot-dashed lines, respectively; the coloured areas in between the lines, yellow and pink respectively, represent the uncertainty region from excluding/including Compton-thick AGN with $25 < \log N_H/{\rm cm}^{-2} < 26$). Irrespective of the input Eddington ratio distributions, both accretion models, which use a radiative efficiency of $\varepsilon = 0.15$, are in good agreement with the unbiased local black hole mass function (cyan region). The grey areas in both panels mark the regions in which local measurements are less robust and cannot be constrained by clustering data.

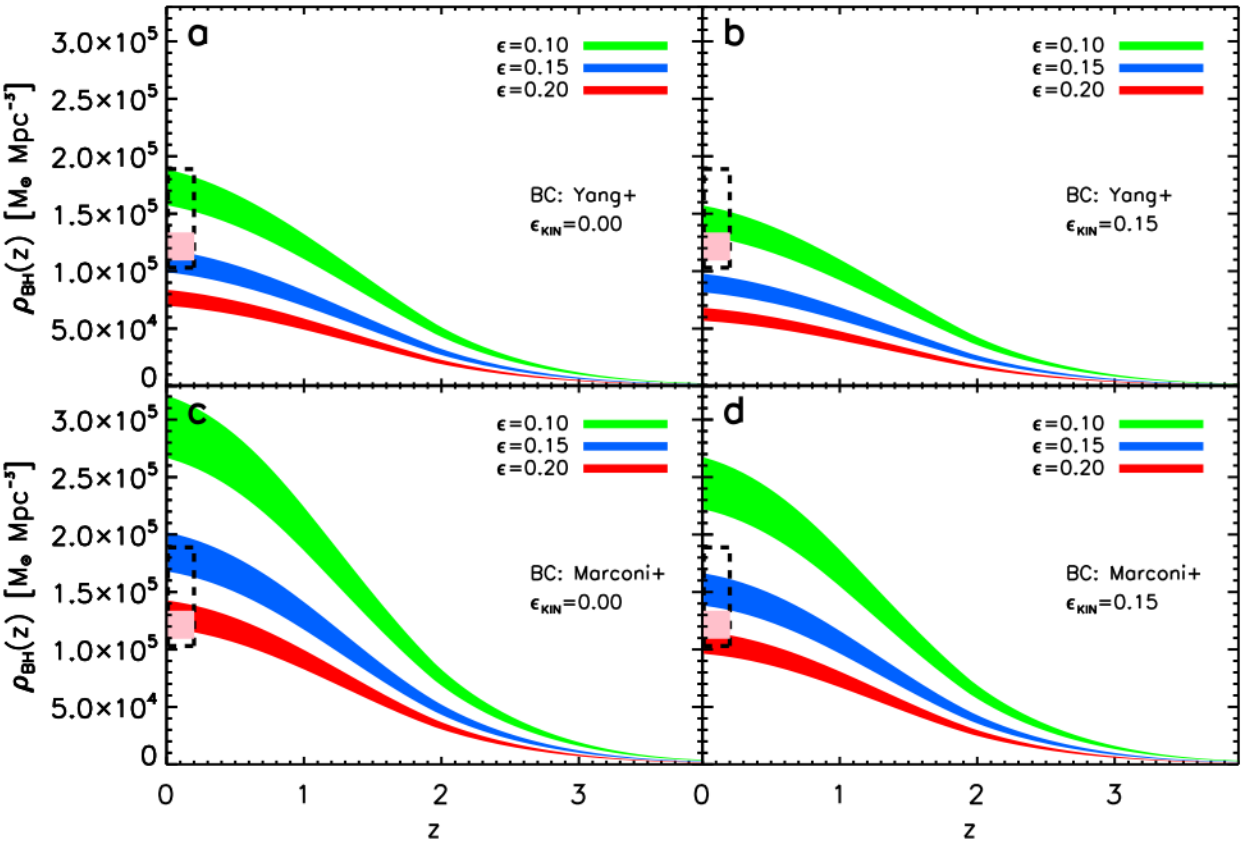

Figure 4: **Comparing local and accreted integrated mass densities.** "Relic" black hole mass density $\rho_{\rm SMBH}$ inferred from the local unbiased[8] $M_{\rm bh}$-$M_{\rm star}$ and $M_{\rm bh}$-σ relations (black dashed box and solid pink square, respectively), compared with the accreted mass density $\rho_{\rm AGN}$ in AGN assuming different values of the radiative efficiency ε, bolometric correction BC, and kinetic efficiency $\varepsilon_{\rm kin}$, as labelled. The coloured region for each model shows the uncertainty from excluding/including Compton-thick AGN with $25 < \log N_H/{\rm cm}^{-2)} < 26$. For the range of currently accepted values of $\varepsilon_{\rm kin}$, BC and obscured fraction, the data indicate moderate-to-high radiative efficiencies ($\varepsilon \gtrsim 0.1$).

the redshift-dependent $M_{\rm bh}$-$M_{\rm star}$ relation from integrating the AGN X-ray emissivity at different epochs along (putative) stellar mass accretion histories. By assuming a constant mean radiative efficiency of ε=0.1 to convert from AGN bolometric luminosities to black hole accretion rates, they claimed a local $M_{\rm bh}$-$M_{\rm star}$ relation in close agreement to the unbiased one by Shankar et al.[8], and definitely steeper than the one from Kormendy & Ho[9].

It is clear from Figure 4 that, irrespective of the exact assumed value of the obscured fraction, bolometric correction or kinetic efficiency, all accretion models always require relatively high radiative efficiencies $\varepsilon \sim 0.10 - 0.20$ to match the local black hole mass density. This yields a mean radiative efficiency a factor of $\sim 2-3$ higher than previous estimates solely based on observed/biased black hole scaling relations[5,36]. Our results would support black hole spin parameters[2] $a \gtrsim 0.7$, which are consistent with a number of direct measurements performed via X-ray reflection analysis[40] or spectral energy distribution fits[41,42]. Hints for relatively high mean radiative efficiencies are also present in some previous works[43,44]. However those studies adopted biased estimates of the local black hole mass density, sometimes **relatively** crude estimates of AGN emissivities, and without the vital constraints from AGN clustering. In this work we have ultimately shown that the AGN large-scale clustering as a function of black hole mass represents a fundamental discriminator for black hole-galaxy scaling relations. Upcoming large-scale AGN surveys such as eROSITA[45], probing millions of AGN, will generate exquisite measurements of AGN clustering, set stringent constraints of the scaling relations down to the lowest accessible black hole masses, and in turn tighter constraints on the radiative and kinetic efficiencies of accreting supermassive black holes.

References

1. Rees, M. J. Black Hole Models for Active Galactic Nuclei. ARA&A **22**, 471–506 (1984).

2. Bardeen, J. M., Press, W. H. & Teukolsky, S. A. Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation. ApJ **178**, 347–370 (1972).

3. Soltan, A. Masses of quasars. MNRAS **200**, 115–122 (1982).

4. Salucci, P., Szuszkiewicz, E., Monaco, P. & Danese, L. Mass function of dormant black holes and the evolution of active galactic nuclei. MNRAS **307**, 637–644 (1999). `arXiv:astro-ph/9811102`.

5. Marconi, A. *et al.* Local supermassive black holes, relics of active galactic nuclei and the X-ray background. MNRAS **351**, 169–185 (2004). `arXiv:astro-ph/0311619`.

6. Shankar, F., Weinberg, D. H. & Miralda-Escudé, J. Accretion-driven evolution of black holes: Eddington ratios, duty cycles and active galaxy fractions. MNRAS **428**, 421–446 (2013). `1111.3574`.

7. Aversa, R., Lapi, A., de Zotti, G., Shankar, F. & Danese, L. Black Hole and Galaxy Coevolution from Continuity Equation and Abundance Matching. ApJ **810**, 74 (2015). 1507.07318.

8. Shankar, F. *et al.* Selection bias in dynamically measured supermassive black hole samples: its consequences and the quest for the most fundamental relation. MNRAS **460**, 3119–3142 (2016). 1603.01276.

9. Kormendy, J. & Ho, L. C. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. ARA&A **51**, 511–653 (2013). 1304.7762.

10. Davis, B. L., Graham, A. W. & Cameron, E. Black Hole Mass Scaling Relations for Spiral Galaxies. II. M $_{BH}$-M $_{*,tot}$ and M $_{BH}$-M $_{*,disk}$. ApJ **869**, 113 (2018). 1810.04888.

11. Busch, G. *et al.* A low-luminosity type-1 QSO sample. I. Overluminous host spheroidals or undermassive black holes. A&A **561**, A140 (2014). 1310.0272.

12. Reines, A. E. & Volonteri, M. Relations between Central Black Hole Mass and Total Galaxy Stellar Mass in the Local Universe. ApJ **813**, 82 (2015). 1508.06274.

13. Shankar, F. *et al.* Black hole scaling relations of active and quiescent galaxies: Addressing selection effects and constraining virial factors. MNRAS (2019). 1901.11036.

14. Bernardi, M., Sheth, R. K., Tundo, E. & Hyde, J. B. Selection Bias in the $M_{\odot}$-σ and $M_{\odot}$-L Correlations and Its Consequences. ApJ **660**, 267–275 (2007). arXiv:astro-ph/0609300.

15. Morabito, L. K. & Dai, X. A Bayesian Monte Carlo Analysis of the M-σ Relation. ApJ **757**, 172 (2012). 1208.3900.

16. Cooray, A. & Sheth, R. Halo models of large scale structure. Phys. Rep. **372**, 1–129 (2002). arXiv:astro-ph/0206508.

17. Shankar, F. *et al.* Revisiting the Bulge-Halo Conspiracy. I. Dependence on Galaxy Properties and Halo Mass. ApJ **840**, 34 (2017). 1703.06145.

18. Grylls, P. J., Shankar, F., Zanisi, L. & Bernardi, M. A statistical semi-empirical model: satellite galaxies in groups and clusters. MNRAS **483**, 2506–2523 (2019). 1812.00015.

19. Davé, R. *et al.* Simba: Cosmological Simulations with Black Hole Growth and Feedback. *arXiv e-prints* (2019). 1901.10203.

20. Savorgnan, G. A. D., Graham, A. W., Marconi, A. & Sani, E. Supermassive Black Holes and Their Host Spheroids. II. The Red and Blue Sequence in the M_{BH}-$M_{*,sph}$ Diagram. ApJ **817**, 21 (2016). 1511.07437.

21. Sahu, N., Graham, A. W. & Davis, B. L. Black Hole Mass Scaling Relations for Early-Type Galaxies: M_BH-M_*, sph and M_BH-M_*, gal. *arXiv e-prints* (2019). 1903.04738.

22. Baron, D. & Ménard, B. Black hole mass estimation for Active Galactic Nuclei from a new angle. *arXiv e-prints* arXiv:1903.01996 (2019). 1903.01996.

23. Powell, M. C. *et al.* The Swift/BAT AGN Spectroscopic Survey. IX. The Clustering Environments of an Unbiased Sample of Local AGNs. ApJ **858**, 110 (2018). 1803.07589.

24. Krumpe, M. *et al.* The Spatial Clustering of ROSAT All-Sky Survey Active Galactic Nuclei. IV. More Massive Black Holes Reside in More Massive Dark Matter Halos. ApJ **815**, 21 (2015). 1509.01261.

25. Krumpe, M., Miyaji, T., Coil, A. L. & Aceves, H. Spatial clustering and halo occupation distribution modelling of local AGN via cross-correlation measurements with 2MASS galaxies. MNRAS **474**, 1773–1786 (2018). 1710.05638.

26. Ghez, A. M. *et al.* Measuring Distance and Properties of the Milky Way's Central Supermassive Black Hole with Stellar Orbits. ApJ **689**, 1044–1062 (2008). 0808.2870.

27. Posti, L. & Helmi, A. Mass and shape of the Milky Way's dark matter halo with globular clusters from Gaia and Hubble. A&A **621**, A56 (2019). 1805.01408.

28. Shankar, F., Bernardi, M. & Sheth, R. K. Selection bias in dynamically-measured supermassive black hole samples: dynamical masses and dependence on Sérsic index. MNRAS (2017). 1701.01732.

29. Sarria, J. E. *et al.* The M_{BH} - M_{star} relation of obscured AGNs at high redshift. A&A **522**, L3 (2010). 1010.0768.

30. Falomo, R., Bettoni, D., Karhunen, K., Kotilainen, J. K. & Uslenghi, M. Low-redshift quasars in the Sloan Digital Sky Survey Stripe 82. The host galaxies. MNRAS **440**, 476–493 (2014). 1402.4300.

31. Tinker, J. *et al.* Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. ApJ **688**, 709–728 (2008). 0803.2706.

32. DiPompeo, M. A., Runnoe, J. C., Hickox, R. C., Myers, A. D. & Geach, J. E. The impact of the dusty torus on obscured quasar halo mass measurements. MNRAS **460**, 175–186 (2016). 1604.06811.

33. Jiang, N. *et al.* Differences in Halo-scale Environments between Type 1 and Type 2 AGNs at Low Redshift. ApJ **832**, 111 (2016). 1602.08825.

34. Shankar, F., Weinberg, D. H. & Shen, Y. Constraints on black hole duty cycles and the black hole-halo relation from SDSS quasar clustering. MNRAS **406**, 1959–1966 (2010). 1004.1173.

35. Sheth, R. K. & Tormen, G. Large-scale bias and the peak background split. MNRAS **308**, 119–126 (1999). arXiv:astro-ph/9901122.

36. Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T. & Watson, M. G. Toward the Standard Population Synthesis Model of the X-Ray Background: Evolution of X-Ray Luminosity and Absorption Functions of Active Galactic Nuclei Including Compton-thick Populations. ApJ **786**, 104 (2014). 1402.1836.

37. Yang, G. *et al.* Linking black hole growth with host galaxies: the accretion-stellar mass relation and its cosmic evolution. MNRAS **475**, 1887–1911 (2018). 1710.09399.

38. Harrison, F. A. *et al.* The NuSTAR Extragalactic Surveys: The Number Counts of Active Galactic Nuclei and the Resolved Fraction of the Cosmic X-Ray Background. ApJ **831**, 185 (2016). 1511.04183.

39. Shankar, F., Cavaliere, A., Cirasuolo, M. & Maraschi, L. Optical-Radio Mapping: the Kinetic Efficiency of Radio-Loud AGNs. ApJ **676**, 131–136 (2008). 0712.3004.

40. Reynolds, C. S. Observing black holes spin. *Nature Astronomy* **3**, 41–47 (2019). 1903.11704.

41. Shankar, F. *et al.* The Optical-UV Emissivity of Quasars: Dependence on Black Hole Mass and Radio Loudness. ApJ **818**, L1 (2016). 1601.02021.

42. Zhang, X. & Lu, Y. On Constraining the Growth History of Massive Black Holes via Their Distribution on the Spin-Mass Plane. *arXiv e-prints* arXiv:1902.07056 (2019). 1902.07056.

43. Elvis, M., Risaliti, G. & Zamorani, G. Most Supermassive Black Holes Must Be Rapidly Rotating. ApJ **565**, L75–L77 (2002). astro-ph/0112413.

44. Yu, Q. & Lu, Y. Toward Precise Constraints on the Growth of Massive Black Holes. ApJ **689**, 732–754 (2008). 0808.3777.

45. Merloni, A. *et al.* eROSITA Science Book: Mapping the Structure of the Energetic Universe. *ArXiv e-prints* (2012). 1209.3114.

Methods

Details on the different $M_{\rm bh}$-$M_{\rm star}$ relations. The left panel of Figure 1 shows a number of linear fits to the $M_{\rm bh}$-$M_{\rm star}$ relation from different groups and black hole samples. Following Shankar et al.[13], we first fitted the subsample of dynamical masses from Savorgnan et al.[20] with reliable black hole mass measurements. We adopted the Bayesian-based linear fitting IDL routine $LINMIXERR$ to compute all median slopes and normalizations. We verified that the variance-based fitting prescriptions reported in Appendix A of Shankar et al.[28] provide identical mean values. The (intrinsic) scatter was instead computed as $\sigma = \sqrt{y_{\rm rms}^2 - \langle\Delta(y_i)\rangle^2}$, where

$y^2_{\rm rms} = \sum_i^N \left[y_i - y(x)\right]^2 / (N-1)$ with $y(x)$ the linear best-fit, and $\Delta(y_i)$ are the quoted uncertainties in black hole masses. The resulting fit to the (full) Savorgnan et al. sample (long-dashed black line) reads as

$$\log \frac{M_{\rm bh}}{M_\odot} = 8.35 \pm 0.09 + (1.31 \pm 0.22) \times \log \left(\frac{M_{\rm star}}{10^{11} M_\odot} \right) . \quad (1)$$

with an intrinsic scatter of 0.5 ± 0.1 dex.

The zero points of the correlations from Davis et al.[10] and Sahu et al.[21], their Equations 3 and 11, respectively, have been modified setting the parameter v in their fits equal to $v = 0.64$, to allow for conversion to the mass-to-light ratios of Bell et al.[46] with a Chabrier[47] stellar initial mass function adopted as a reference in Shankar et al. (2016)[8] and in this work. The Reines & Volonteri fit[12] and the Kormendy & Ho[9] sample, which is also taken from Table 3 of Reines & Volonteri, have been corrected for the different mass-to-light ratios following Equation A1 in Shankar et al. (2019)[13]. A direct linear fit to the (corrected) Kormendy & Ho[9] sample retrieved from Reines & Volonteri yields $\log \frac{M_{\rm bh}}{M_\odot} = 8.56 + 1.58 \times \log \left(\frac{M_{\rm star}}{10^{11} M_\odot} \right)$ (dot-dashed cyan line in the left panel of Figure 1). The sample by Baron & Ménard[22] has been converted to the Bell et al. mass-to-light ratios adopting the mass-dependent correction presented in Figure A2 of Bernardi et al. (2017)[48].

Last but not least, in the same left panel of Figure 1 we report the intrinsic (or unbiased/debiased) correlation put forward by Shankar et al. (2016)[8], which is the result of a series of Monte Carlo simulations. The mean intrinsic relation by Shankar et al. between black hole mass and total host galaxy stellar mass $M_{\rm bh}$-$M_{\rm star}$ is well approximated by the relation

$$\log \frac{M_{\rm bh}}{M_\odot} = 7.574 + 1.946 \, \log \left(\frac{M_{\rm star}}{10^{11} M_\odot} \right) - 0.306 \times \left[\log \left(\frac{M_{\rm star}}{10^{11} M_\odot} \right) \right]^2 - 0.011 \, \left[\log \left(\frac{M_{\rm star}}{10^{11} M_\odot} \right) \right]^3 , \quad (2)$$

with a mass-dependent intrinsic scatter in (the logarithm of) black hole mass at fixed stellar mass (yellow region) well reproduced by[49]

$$\Delta \log \frac{M_{\rm bh}}{M_\odot} = 0.32 - 0.1 \times \log \left(\frac{M_{\rm star}}{10^{12} M_\odot} \right) . \quad (3)$$

Both Equations 2 and 3 are good fits[8,13] to galaxies with stellar mass above a few times $\log M_{\rm star}/M_\odot > 10$. As recently discussed by Shankar et al. (2019)[13], Equation 2 was obtained by imposing that black holes closely follow, via the $M_{\rm bh} \propto \sigma^{4-5}$ relation, the velocity dispersion-stellar mass of SDSS early-type galaxies. Any significant contribution from later-type galaxies would tend, if anything, to produce slightly lower normalizations of the unbiased $M_{\rm bh}$-$M_{\rm star}$, thus further increasing the mismatch with the scaling relation of quiescent galaxies.

Abundance matching technique. The $M_{\rm star}$-$M_{\rm halo}$ relation is obtained from abundance matching between the stellar mass and the halo plus subhalo mass functions[17,50]. We here adopt the Tinker et al. (2008) halo mass function[31] and the unevolved, surviving subhalo mass function by Jiang and van den Bosch[51] (see their Section 4.5). The latter is the relevant statistical description of subhaloes that survived mergers down to a given redshift, and thus the most suited quantity to include in accurate abundance matching algorithms. However, we verified that adopting different, more general fits of the subhalo mass function[52] would yield very similar results.

We assume throughout an intrinsic scatter of 0.15 dex in stellar mass at fixed halo mass[17]. We include scatter following Equation 5 in Aversa et al.[7], though very similar results are found by following the methodology presented in Shankar et al. (2017)[17]. We adopt the Sérsic+Exponential stellar mass function from Bernardi et al. 2013[53], which is based on the same photometry and mass-to-light ratios used by Shankar et al. (2016) and in this work. We assume no evolution in the stellar mass function up to $z = 0.25$, which is a very good approximation according to the latest estimates[54,55].

The $M_{\rm star}$-$M_{\rm halo}$ relation is now more securely established at least for relatively high stellar masses $M_{\rm star} \gtrsim 3 \times 10^{10}\,M_{\odot}$ and at relatively low redshifts[17,18,53,56–58]. We then extract large host halo catalogues from the halo plus subhalo mass function and assign galaxies to haloes via the $M_{\rm star}$-$M_{\rm halo}$ relation. We note that all central and satellite galaxies are here assigned from the $M_{\rm star}$-$M_{\rm halo}$ relation at the single redshift of $z = 0.25$, which is an extremely good approximation to a full multi-epoch abundance matching routine, especially for relatively massive galaxies ($M_{\rm star} \gtrsim 3 \times 10^{10}, M_{\odot}$), as recently demonstrated by Grylls et al.[18]. In this work we always assume, unless otherwise noted, a standard reference cosmology of $\Omega_m = 0.3$, $h = 0.7$, $\sigma_8 = 0.8$, $n_s = 1$, as adopted in the observational papers we compare with.

Computing local and accreted mass functions. The local black hole mass function is computed from the SDSS galaxy sample adopted by Shankar et al.[8]. In a Monte Carlo fashion, to each galaxy a supermassive black hole is assigned following a given input $M_{\rm bh}$-$M_{\rm star}$ relation. To allow for moderate systematics in the galaxy stellar mass estimates due to fine details in the light profiles and/or mass-to-light ratios, at each iteration we randomly add ± 0.1 dex to the stellar masses of each host galaxy, and then assign to it a supermassive black hole from the $M_{\rm bh}$-$M_{\rm star}$ relation inclusive of its scatter. At each iteration, the black hole mass function is then computed from the $V_{\rm max}$ weights assigned to each SDSS galaxy. The final 1σ error in number density at fixed black hole mass is then given by the dispersion in number densities resulting from 100 realizations. Supplementary Table 1 reports the results of the black hole mass function and its 1σ uncertainty at fixed black hole mass derived from the $M_{\rm bh}$-$M_{\rm star}$ relation. The black hole mass function implied by the $M_{\rm bh}$-σ relation is fully consistent with the one given in Supplementary Table 1 within the uncertainties, as shown in Figure 3.

In Supplementary Figure 5 we show the results of applying the Monte Carlo methodology described above to only elliptical galaxies, which we select from the SDSS sample imposing a Bayesian classification probability[59] $P[E] > 0.5$ (our results do not depend on the exact value

of this threshold). The left panel shows the comparison between the black hole mass functions implied by the observed $M_{\rm bh}$-$M_{\rm star}$ and $M_{\rm bh}$-σ relations of early-type galaxies[13]. A clear mismatch of the order of $\sim 2-3$ in integrated black hole mass density, is evident between the black hole mass functions, proving that the difference in black hole mass estimators is not a simple byproduct of including or not bulge corrections in the calculation. On the other hand, the right panel of Figure 5 shows the good match between the black hole mass functions implied by the unbiased/intrinsic $M_{\rm bh}$-$M_{\rm star}$ and $M_{\rm bh}$-σ relations of early-type galaxies.

To compute the accreted (or "relic") black hole mass function we instead solve the continuity equation[60–62]

$$\frac{\partial n_{\rm bh}}{\partial t}(M_{\rm bh},t) = -\frac{\partial(\langle \dot{M}_{\rm bh}\rangle n_{\rm bh}(M_{\rm bh},t))}{\partial M_{\rm bh}}\,, \tag{4}$$

where $\langle \dot{M}_{\rm bh}\rangle$ is the mean accretion rate (averaged over the active and inactive populations) of the black holes of mass $M_{\rm bh}$ and number density $n_{\rm bh}(M_{\rm bh},t)$ at time t.

In the main text we show results for the accreted black hole mass function computed in two ways, via the method outlined in Shankar et al.[6] and as proposed by Aversa et al.[7]. In the former, Equation 4 is solved by assuming[63] an input Eddington ratio distribution (lognormal) and constantly peaked around an Eddington ratio $\log\lambda = -0.6$, with $\lambda \propto L/M_{\rm bh}$) of active black holes convolved with the active mass function to yield the observed AGN luminosity function at any cosmic epoch. In the latter, following the seminal work by Yu & Lu[64], a light curve is instead assumed in input and the active and total black hole mass functions are then self-consistently derived by direct analytic time integration. Aversa et al. also adopted a slim and thin-disk approximation, in which the radiative efficiency is assumed to depend on the current Eddington ratio, in which the radiative efficiency can be close to 0.3 during super-Eddington accretion, but rapidly approaching the thin-disc constant value at Eddington/sub-Eddington regimes. Given that the latter is the most typical accretion mode for AGN, the Aversa et al. model reduces to a standard, constant radiative efficiency throughout most of the evolution time.

The normalization and shape of the implied duty cycle, i.e., the ratio between the active and total black hole mass functions, depend on the input Eddington ratio distributions $P(\lambda)$, with $\lambda \propto L/M_{\rm bh}$. As extensively discussed by Shankar et al.[6], assuming, for example, a constant Eddington ratio distribution, gradually generates a very low duty cycle at fixed black hole mass at late epochs, which is largely inconsistent with present data[65]. On the other hand, an Eddington ratio distribution that drops towards lower values of $\lambda(z)$ at lower redshifts, can provide orders of magnitude higher duty cycles, as the same black hole mass bins are mapped to lower luminosity and more abundant AGN. We checked that increasing the radiative efficiency from $\varepsilon \sim 0.05$ to $\varepsilon \sim 0.2$ does not change these behaviours in the output duty cycle.

Accreted mass densities. The classical[3,4,66] argument based on integrated mass densities, and inclusive of both radiative and kinetic efficiencies[39,67,68], can be cast as follows

$$\rho_{\rm AGN} = \int dz \int dL \Phi(L,z) L \frac{(1-\varepsilon-\varepsilon_{\rm kin})}{\varepsilon c^2} = \rho_{\rm SMBH}\,, \tag{5}$$

in which the integrated (or "relic") mass density $\rho_{\rm AGN}$ from accretion over all AGN with number densities $\Phi(L,z)$, bolometric luminosity L and redshift z, is compared to the local black hole mass density $\rho_{\rm SMBH}$. The integral in Equation 5 is computed over the luminosity range $42 < \log L/{\rm erg\,s^{-1}} < 48$ and from $z = 6$. The exact limits of integration do not significantly change the results. The values of bolometric luminosity L in input into Equation 5, implicitly assume a bolometric correction $k_{\rm bol}(L)$ that is luminosity dependent, folded in the computation of the bolometric luminosity function $\Phi(L,z)$ derived from Ueda et al.'s X-ray luminosity function.

Computing large-scale bias To compare models to clustering data, we first perform abundance matching at the average redshift of $z = 0.25$. We then extract large host halo catalogues from the halo (plus subhalo) mass function and assign galaxies to haloes via the $M_{\rm star}$-$M_{\rm halo}$ relation as described above.

On the general assumption that black hole scaling relations do not depend on environment/host halo mass, to each mock galaxy in the catalogue we assign a black hole mass adopting an input $M_{\rm bh}$-$M_{\rm star}$ relation inclusive of scatter. In principle, the mean large-scale bias for black holes with mass in the range $M_{\rm bh}$ and $M_{\rm bh} + dM_{\rm bh}$ could be straightforwardly estimated as

$$b(M_{\rm bh}) = \frac{1}{N_{\rm bin}} \sum_{i=1}^{N_{\rm bin}} b_h \left[M_{\rm halo,i}(M_{\rm bh})\right] \,, \tag{6}$$

where the sum runs over the total number $N_{\rm bin}$ of (central and satellite) parent haloes hosting central black holes with mass in the range $M_{\rm bh}$ and $M_{\rm bh} + dM_{\rm bh}$.

However, Equation 6 neglects the probabilities for central and satellite black holes to be active. In the case of AGN in fact, not all galaxies in a given host halo mass bin necessarily contribute to the same clustering signal, and Equation 6 should be modified to include the duty cycles, or probabilities, $U_{\rm cen}(M_{\rm bh})$ and $U_{\rm sat}(M_{\rm bh})$ of, respectively, central and satellite black holes of mass $M_{\rm bh}$ to be active at a given cosmic epoch. We here follow the same convention as in Shankar et al. (2013)[6] and denote the duty cycles at fixed black hole mass $M_{\rm bh}$ at a given redshift z as $U(M_{\rm bh}, z)$. As in this work we only focus on the specific redshift of $z = 0.25$, we will drop the redshift dependence in the duty cycles from here onwards. Generalising Equation 6 to include the central/satellite probabilities of being active we obtain

$$b(M_{\rm bh}) = \frac{\left[\sum_{i=1}^{N_{\rm cen}(M_{\rm bh})} U_{\rm cen,i}(M_{\rm bh})(M_{\rm bh}) b_{\rm cen,i}(M_{\rm bh}) + \sum_{i=1}^{N_{\rm sat}(M_{\rm bh})} U_{\rm sat,i}(M_{\rm bh}) b_{\rm sat,i}(M_{\rm bh})\right]}{\left[\sum_{i=1}^{N_{\rm cen}(M_{\rm bh})} U_{\rm cen,i}(M_{\rm bh}) + \sum_{i=1}^{N_{\rm sat}(M_{\rm bh})} U_{\rm sat,i}(M_{\rm bh})\right]} \,, \tag{7}$$

where both $U_{\rm cen}(M_{\rm bh})$ and $U_{\rm sat}(M_{\rm bh})$ are related to the total duty cycle of active black holes by the relation $N_{\rm AGN}(M_{\rm bh}) = U(M_{\rm bh})N(M_{\rm bh}) = U_{\rm cen}(M_{\rm bh})N_{\rm cen}(M_{\rm bh}) + U_{\rm sat}(M_{\rm bh})N_{\rm sat}(M_{\rm bh})$, with $N(M_{\rm bh}) = N_{\rm cen}(M_{\rm bh}) + N_{\rm sat}(M_{\rm bh})$ the number of central and satellite black holes in the mass bin $M_{\rm bh}$ and $M_{\rm bh} + dM_{\rm bh}$. In the limit in which all central and satellite black holes are active or share equal probabilities to be active, i.e., $U_{\rm cen}(M_{\rm bh}) = U_{\rm sat}(M_{\rm bh})$, then Equation 7 reduces to the special case of Equation 6.

Throughout this work we adopt as a reference the total duty cycle $U(M_{\rm bh})$ in Equation 7 as derived from the continuity equation model by Shankar et al. (2013)[6], with constant lognormal Eddington ratio distribution. The exact normalization and shape of the total duty cycle $U(M_{\rm bh})$ depends on the minimum luminosity threshold considered and input Eddington ratio distributions $P(\lambda)$, as anticipated in the previous Section . As extensively discussed by, e.g., Shankar et al.[6], continuity equation models generally tend to generate duty cycles decreasing with increasing black hole mass, as also retrieved by direct data modelling by other groups[6]. The latter trend would then imply that active lower mass black holes would have a larger weight on the bias (as in Figure 10 discussed below). However, when computing the mean bias in narrow bins of black hole mass (as in Figure 2), the exact shape or normalization of the assumed total duty cycle $U(M_{\rm bh})$ are irrelevant. It is in fact clear from Equation 7 that, in a relatively small bin of black hole mass, what contributes to the mean bias is not the total duty cycle $U(M_{\rm bh})$, rather the *relative* probabilities for central and satellite black holes to be active, i.e., the ratio $Q_{\rm bh}(M_{\rm bh}) = U_{\rm sat}(M_{\rm bh})/U_{\rm cen}(M_{\rm bh})$.

Although satellite black holes are usually relatively less abundant than central black holes of similar mass, especially at progressively larger masses of interest to this work, if the ratio $Q_{\rm bh} \gtrsim 1$, satellite black holes could still noticeably contribute to the total mean bias as they inhabit very massive host dark matter haloes. Several works in the literature[69–72] have constrained the relative fraction of active black holes residing in satellite haloes to be around $f_{\rm sat}^{\rm AGN} \lesssim 0.1 - 0.2$. In terms of duty cycles, at any given black hole mass, the AGN satellite fraction $f_{\rm sat}^{\rm AGN}$ would translate into

$$f_{\rm sat}^{\rm AGN}(M_{\rm bh}) = \frac{U_{\rm sat}(M_{\rm bh})N_{\rm sat}(M_{\rm bh})}{U(M_{\rm bh})N(M_{\rm bh})} = \frac{[Q_{\rm bh}(M_{\rm bh})N_{\rm sat}(M_{\rm bh})/N_{\rm cen}(M_{\rm bh})]}{[Q_{\rm bh}(M_{\rm bh})N_{\rm sat}(M_{\rm bh})/N_{\rm cen}(M_{\rm bh}) + 1]}\,, \tag{8}$$

implying

$$Q_{\rm bh}(M_{\rm bh}) = \frac{f_{\rm sat}^{\rm AGN}(M_{\rm bh})}{1 - f_{\rm sat}^{\rm AGN}(M_{\rm bh})}\frac{1 - f_{\rm sat}^{\rm BH}}{f_{\rm sat}^{\rm BH}}\,, \tag{9}$$

where $f_{\rm sat}^{\rm BH}(M_{\rm bh}) = N_{\rm sat}(M_{\rm bh})/[N_{\rm sat}(M_{\rm bh}) + N_{\rm cen}(M_{\rm bh})]$ is the total fraction of satellite black holes with mass within $M_{\rm bh}$ and $M_{\rm bh}+dM_{\rm bh}$. Based on our mocks, the total fraction of satellites for the high-mass black holes considered by Krumpe et al.[24], is $f_{\rm sat}^{\rm BH}(M_{\rm bh}) \sim 0.1$, which corresponds to $Q_{\rm bh}(M_{\rm bh}) \sim 2$ when adopting the best-fit $f_{\rm sat}^{\rm AGN} = 0.18$ from Leauthaud et al.[71]. We will always adopt the latter value of $Q_{\rm bh} = 2$ independent of black hole mass as our reference value throughout this work, though we will also show results with $Q_{\rm bh} = 1$. Indeed, $Q_{\rm bh}$ could take on even lower values. Other groups measured a $f_{\rm sat}^{\rm AGN}$ consistent with zero[69]. More recently Man et al.[73] found the SDSS completeness-corrected fraction of local AGN at fixed stellar mass is independent of environment, with the fraction of AGN in satellite galaxies being comparable, if not lower, than that in central galaxies (see their Figure 4). The Man et al. results would imply $Q_{\rm bh} \lesssim 1$ largely independent of host stellar mass, as assumed here.

When computing the mean bias in Figure 2 we make use of Equation 7 with $Q_{\rm bh} = 2$ and the halo bias b_h is taken from Tinker et al. (2005)[74], in line with the Halo Occupation Distribution models adopted by Krumpe et al.[24]. We note that Tinker et al. make use of FoF-based halo masses. We thus correct our reference halo masses, defined as 200 times the critical density, by an average

factor[75] of 0.966 before applying the Tinker et al. analytic model, though this has a very minor effect on the results. For completeness, we proceed below to a more thorough comparison with the Krumpe et al.[24] full cross-correlation function between RASS (SDSS) AGN and the Luminous Red Galaxies (LRGs), as shown with filled black squares in Figure 7 for the SDSS Data Release 4 (top panels) and Data Release 7 (bottom panels). The (projected) cross-correlation function at large scales is analytically computed at any redshift of interest z as

$$w_P(r_P, z) = b_{\rm AGN} b_{\rm LRG} w_{\rm DM}(r_P, z)\,, \tag{10}$$

where $b_{\rm AGN}$ and $b_{\rm LRG}$ are the median large scale biases of AGN and LRGs, respectively, and $w_{\rm DM}(r_P, z)$ is the projected linear matter correlation function derived from the linear matter power spectrum of Smith et al. (2003)[76], though we verified that other power spectra yield nearly identical results.

The bias of LRGs is computed by fitting the predicted linear auto-correlation function

$$w_P(r_P, z) = b^2_{\rm LRG} w_{\rm DM}(r_P, z)\,, \tag{11}$$

to the data. Including the full covariance matrix of the uncertainty on the pair counts, a direct χ^2 fitting[77] yields $b_{\rm LRG} = 2.25 \pm 0.03$ for both DR4 and DR7. The covariance matrix was estimated in Krumpe et al.[24] (see their Equation 6) by using jackknife resampling, dividing the survey area in 100 subsection (131) for X-ray (optically) selected AGN. The estimated value of the bias for the LRGs would imply an effective host halo mass of $M_{\rm 200c} \sim 6 \times 10^{13}\,M_\odot/h$ and a mean stellar mass of $M_{\rm star} \sim 5 \times 10^{11}\,M_\odot$, in good agreement with the independent analysis by van Uitert et al. 2015[78] on the LOWZ and CMASS LRGs from the Baryon Oscillation Spectroscopic Survey.

Fixing $b_{\rm LRG} = 2.25$ in Equation 10, we find, when integrating over a scale of $3-30$ Mpc and including the full covariance in the χ^2 fitting, $b_{\rm AGN} = 1.43 \pm 0.11$ and $b_{\rm AGN} = 1.56 \pm 0.10$ for DR7 and DR4, respectively. In our mocks we then select a black hole mass distribution consistent with the one adopted by Krumpe et al., i.e., with a lognormal distribution peaked at $\log M_{\rm bh}/M_\odot = 8.5$ and with a dispersion of 0.3 dex (the "high $M_{\rm BH}$ sample" in their Figure 5). We then compute the mean bias as given in Equation 7 (we stress that medians would, if anything, tend to produce lower biases) assuming $Q_{\rm bh} = 1$ (left panels) and $Q_{\rm bh} = 2$ (right panels), which would correspond to a fraction $f_{\rm sat} \sim 0.1-0.2$, following Equation 9. When computing the full cross-correlation function we use the more recent analytic model for the bias put forward by Tinker et al. (2010)[79], which provides a closer match to the halo autocorrelation function from the MultiDark simulation[80], as shown in Supplementary Figure 6. Keeping the Tinker et al. (2005) bias would, if anything, further strengthen our conclusion that the clustering data are most consistent with the intrinsic/unbiased black hole scaling relations.

Equation 7 yields a mean $b_{\rm AGN} = 1.39$ and $b_{\rm AGN} = 1.08$, for $Q_{\rm bh} = 1$ and $b_{\rm AGN} = 1.42$ and $b_{\rm AGN} = 1.13$, for $Q_{\rm bh} = 2$, when adopting the intrinsic and observed $M_{\rm bh}$-$M_{\rm star}$ relations, Equations 2 and 1, respectively. The latter model is always $\gtrsim 3 - 3.5\sigma$ away from the best-fit value of the DR7 (and even more for the DR4), in line with the results presented in Figure 2. We

note that all our assumptions so far have been as conservative as possible, aimed at minimizing the tension in the large-scale clustering between the data and the model based on the observed $M_{\rm bh}$-$M_{\rm star}$ relation. Adopting even lower $Q_{\rm bh}$ values, and/or more up-to-date observed $M_{\rm bh}$-$M_{\rm star}$ relations[10], and/or empirically-based duty cycles[81], would all tend to yield mean values of the predicted $b_{\rm AGN}$ close to unity, implying $> 3-4\sigma$ departures from the mean bias directly retrieved from the projected correlation function. The implied cross-correlation functions from the intrinsic and observed models, found by setting $b_{\rm LRG} = 2.25$ into Equation 10, are reported in the Supplementary Figure 7 with solid red and long-dashed black lines, respectively.

Systematics in mean radiative efficiency and obscured fractions This work puts forward a novel methodology that makes use of AGN clustering measurements to set firmer constraints on the shape of the local $M_{\rm bh}$-$M_{\rm star}$ relation and, in turn, on the mean radiative efficiency of AGN. Some key points should be emphasized. When comparing to a given rendition of the local $M_{\rm bh}$-$M_{\rm star}$ relation, the accretion models are actually able to constrain the *ratio* between bolometric correction and radiative efficiency $k_{\rm bol}(1-\varepsilon)/\varepsilon$ (see Equation 5).

We showed in Figure 4 that, when adopting the latest renditions of the (luminosity-dependent) bolometric corrections by Yang et al.[37,82], we require mean radiative efficiencies of the order of $\varepsilon \gtrsim 0.1-0.2$ to match the local black hole mass density implied by the intrinsic/unbiased $M_{\rm bh}$-$M_{\rm star}$ scaling relation[8]. Up to one order of magnitude lower mean radiative efficiencies $\varepsilon \sim 0.02$, would instead be necessary to broadly align with the local black hole mass density inferred from the *observed* $M_{\rm bh}$-$M_{\rm star}$ relation (yellow region in the right panel of Figure 3).

Ueda et al.[36] performed very similar calculations to ours and with the same AGN luminosity function. When adopting as a local reference a black hole mass function derived from the Kormendy & Ho $M_{\rm bh}$-$M_{\rm star}$ scaling relation, comparable to the dot-dashed black lines in Figure 3 (filled red circles in their Figure 23), they derived a best-fit radiative efficiency of $\varepsilon = 0.053$. Correcting their estimate of the mean radiative efficiency based on a different bolometric correction[83], which is a factor of $\sim 2-3$ higher than our reference one, would imply extremely low values of $\varepsilon \lesssim 0.03$, fully consistent with our "biased" estimate of the radiative efficiency. The previous results by Ueda et al. and other works would thus be in tension with estimates of independent spectral estimates of the spin of supermassive black holes[40], which suggest radiative efficiencies of the order of $\varepsilon \gtrsim 0.1$ or above.

Higher values of the bolometric corrections would in fact require proportionally higher radiative efficiencies to generate the same relic black hole mass density (Equation 5). Indeed, Zhang et al.[84] recently retrieved a mean radiative efficiency of $\varepsilon \sim 0.3$ when adopting Shankar et al.'s intrinsic local black hole relations, which is in line with our estimates when considering their $\sim 2-3$ larger values of the radiative efficiency[83,85].

Shankar et al. (2013)[6] also considered alternative accretion models that made use of either an Eddington ratio-dependent bolometric correction from Vasudevan & Fabian[86], or a redshift-dependent mean radiative efficiency. The former model produced a relic black hole mass function

similar in shape to the one based on luminosity-dependent bolometric corrections (see their Figure 14), though with noticeably less massive black holes (their Figure 15), implying an overall lower integrated black hole mass density falling in between the models presented in the top and bottom panels of Figure 4. A mean radiative efficiency decreasing with cosmic time from close to maximal ($\varepsilon \sim 0.2-0.3$) to minimal values ($\varepsilon \sim 0.05$) at low redshift, was originally considered by Shankar et al. as a possibility to reconcile the preliminary analysis of high-redshift clustering data of luminous quasars, which already suggested relatively high radiative efficiencies[87], with the "biased" local black hole mass densities. A progressively decreasing radiative efficiency at late times is however disfavoured by our current results as it would generate relic black hole mass densities too large with respect to our updated local black hole mass densities.

Another important point to mention, which can be of relevance for all statistical supermassive black hole accretion models, is the possible presence of a selection effect on black hole spin distribution affecting flux-limited AGN surveys[88]. Assuming an intrinsically broad spin distribution at, say, fixed black hole mass, flux-limit effects would then favour the detection of the more luminous, higher-spin/higher-ε sources. This in turn would imply that the average intrinsic radiative efficiency could be lower at fixed black hole or host galaxy stellar mass. On the other hand we note that even assuming a minimum of $\varepsilon = 0.05$, setting[89] $L = f_{0.1}\dot{m}lM_{\rm bh}$, with $f = \varepsilon/(1-\varepsilon)$, $l \sim 1.3 \times 10^{38}\,{\rm erg\,s^{-1}\,M_{\odot}^{-1}}$, $\dot{m} = \dot{M}_{\rm bh}/\dot{M}_{\rm Edd}$, and $\dot{M}_{\rm Edd} = L_{\rm Edd}/0.1c^2$, we would get a bolometric luminosity of $L \gtrsim 10^{43}\,{\rm erg\,s^{-1}}$, for $M_{\rm bh} \gtrsim 10^6\,M_{\odot}$ accreting at an average[6] $\dot{m} \sim 0.1$. This limit in bolometric luminosity would correspond to $L_X \gtrsim 10^{42}\,{\rm erg\,s^{-1}}$, for typical X-ray bolometric corrections at these luminosities, which is still within the reach of the surveys adopted in this work, at least at $z \lesssim 2$ (see left panel of Figure 3 in Ueda et al.[36]).

Last but not least, a fundamental source of systematic uncertainty in the calculation of the mean radiative efficiency derived from Equation 5, originates from the fraction of obscured AGN included in the AGN luminosity function. Equation 5 in fact implies specifying a *complete* census of AGN, and thus the AGN luminosity function must be corrected for the fraction of obscured sources $f_{\rm obsc}(L,z)$ missed even in hard X-ray AGN surveys. In this work we followed Ueda et al.[36], who suggest a fraction of obscured Compton-thick AGN with column densities in the range $24 < \log N_H/{\rm cm^{-2}} < 26$ equal to 50% the number densities of AGN with $20 < \log N_H/{\rm cm^{-2}} < 24$. The Ueda et al. estimate is in good agreement with the more recent extrapolated estimates from NuSTAR observations of heavily obscured Swift/BAT AGN by Georgantopoulos and Akylas[90], who predict, from analyzing a fraction of the BAT sample, *up to* 50% of Compton-thick AGN with respect to the total population, within typical AGN spectral parameters. On the other hand, from cutting-edge comprehensive AGN population synthesis models, Ananna et al.[91] have recently claimed a fraction of Compton-thick AGN with $24 < \log N_H/{\rm cm^{-2}} < 26$ a factor of $\sim 2-3$ higher than what estimated by Ueda et al. (their Figure 10). Increasing the number densities of AGN increases the normalization of the AGN luminosity function $\Phi(L,z)$ in Equation 5, implying a higher relic mass density at fixed bolometric correction, thus proportionally higher radiative efficiencies to match the same local black hole mass density.

We note that in hierarchical models of structure formation[75], a substantial population of won-

dering black holes may end up co-existing with the central black hole within the same host galaxy. Accounting for this additional, at present unknown, population of wondering black holes in the estimate of the local black hole mass density would imply a decrease in the mean radiative efficiency if the wondering black holes were themselves active and recorded in the AGN luminosity function. More generally, orbiting and/or ejected black holes could affect the applicability of Equation 5. Cosmological hydrodynamic simulations[92] showed, however, that the correction induced by the integrated contribution of black holes today not locked up at the centre of galaxies, should not amount to more than $\sim 11\%$, on average.

All in all, our inferred mean radiative efficiencies of $\varepsilon \gtrsim 0.1 - 0.2$ can be safely considered as lower limits at fixed kinetic efficiency $\varepsilon_{\rm kin}$ and local black hole mass density, as our adopted bolometric corrections and obscured fractions are comparable to, or if anything lower than, other estimates put forward in the literature.

46. Bell, E. F., McIntosh, D. H., Katz, N. & Weinberg, M. D. The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. ApJS **149**, 289–312 (2003). astro-ph/0302543.

47. Chabrier, G. Galactic Stellar and Substellar Initial Mass Function. PASP **115**, 763–795 (2003). arXiv:astro-ph/0304382.

48. Bernardi, M. *et al.* The high mass end of the stellar mass function: Dependence on stellar population models and agreement between fits to the light profile. MNRAS (2017). 1604.01036.

49. Sesana, A., Shankar, F., Bernardi, M. & Sheth, R. K. Selection bias in dynamically measured supermassive black hole samples: consequences for pulsar timing arrays. MNRAS **463**, L6–L11 (2016). 1603.09348.

50. Shankar, F. *et al.* Revisiting the bulge-halo conspiracy - II. Towards explaining its puzzling dependence on redshift. MNRAS **475**, 2878–2890 (2018). 1711.07986.

51. Jiang, F. & van den Bosch, F. C. Statistics of dark matter substructure - I. Model and universal fitting functions. MNRAS **458**, 2848–2869 (2016).

52. Giocoli, C., Tormen, G. & van den Bosch, F. C. The population of dark matter subhaloes: mass functions and average mass-loss rates. MNRAS **386**, 2135–2144 (2008). 0712.1563.

53. Bernardi, M. *et al.* The massive end of the luminosity and stellar mass functions: dependence on the fit to the light profile. MNRAS **436**, 697–704 (2013). 1304.7778.

54. Bernardi, M. *et al.* The massive end of the luminosity and stellar mass functions and clustering from CMASS to SDSS: evidence for and against passive evolution. MNRAS **455**, 4122–4135 (2016). 1510.07702.

55. Tinker, J. L. *et al.* The Correlation between Halo Mass and Stellar Mass for the Most Massive Galaxies in the Universe. ApJ **839**, 121 (2017). 1607.04678.

56. Kravtsov, A. V., Vikhlinin, A. A. & Meshcheryakov, A. V. Stellar Mass-Halo Mass Relation and Star Formation Efficiency in High-Mass Halos. *Astronomy Letters* **44**, 8–34 (2018). 1401.7329.

57. Behroozi, P., Wechsler, R., Hearin, A. & Conroy, C. UniverseMachine: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from z=0-10. *ArXiv e-prints* (2018). 1806.07893.

58. Moster, B. P., Naab, T. & White, S. D. M. EMERGE - an empirical model for the formation of galaxies since z 10. MNRAS **477**, 1822–1852 (2018). 1705.05373.

59. Huertas-Company, M., Aguerri, J. A. L., Bernardi, M., Mei, S. & Sánchez Almeida, J. Revisiting the Hubble sequence in the SDSS DR7 spectroscopic sample: a publicly available Bayesian automated classification. A&A **525**, A157 (2011). 1010.3018.

60. Small, T. A. & Blandford, R. D. Quasar evolution and the growth of black holes. MNRAS **259**, 725–737 (1992).

61. Yu, Q. & Tremaine, S. Observational constraints on growth of massive black holes. MNRAS **335**, 965–976 (2002). astro-ph/0203082.

62. Shankar, F., Salucci, P., Granato, G. L., De Zotti, G. & Danese, L. Supermassive black hole demography: the match between the local and accreted mass functions. MNRAS **354**, 1020–1030 (2004). arXiv:astro-ph/0405585.

63. Cao, X. Cosmological Evolution of Massive Black Holes: Effects of Eddington Ratio Distribution and Quasar Lifetime. ApJ **725**, 388–393 (2010). 1010.0046.

64. Yu, Q. & Lu, Y. Constraints on QSO Models from a Relation between the QSO Luminosity Function and the Local Black Hole Mass Function. ApJ **602**, 603–624 (2004). arXiv:astro-ph/0311404.

65. Goulding, A. D., Alexander, D. M., Lehmer, B. D. & Mullaney, J. R. Towards a complete census of active galactic nuclei in nearby galaxies: the incidence of growing black holes. MNRAS **406**, 597–611 (2010). 1003.3015.

66. Shankar, F. Black hole demography: from scaling relations to models. *Classical and Quantum Gravity* **30**, 244001 (2013). 1307.3289.

67. Ghisellini, G., Haardt, F., Della Ceca, R., Volonteri, M. & Sbarrato, T. The role of relativistic jets in the heaviest and most active supermassive black holes at high redshift. MNRAS **432**, 2818–2823 (2013). 1304.1152.

68. Zubovas, K. AGN must be very efficient at powering outflows. MNRAS **479**, 3189–3196 (2018). 1806.08914.

69. Starikova, S. *et al.* Constraining Halo Occupation Properties of X-Ray Active Galactic Nuclei Using Clustering of Chandra Sources in the Boötes Survey Region. ApJ **741**, 15 (2011). 1010.1577.

70. Shen, Y. *et al.* Cross-correlation of SDSS DR7 Quasars and DR10 BOSS Galaxies: The Weak Luminosity Dependence of Quasar Clustering at z ˜ 0.5. ApJ **778**, 98 (2013). 1212.4526.

71. Leauthaud, A. *et al.* The dark matter haloes of moderate luminosity X-ray AGN as determined from weak gravitational lensing and host stellar masses. MNRAS **446**, 1874–1888 (2015). 1410.5817.

72. Rodríguez-Torres, S. A. *et al.* Clustering of quasars in the first year of the SDSS-IV eBOSS survey: interpretation and halo occupation distribution. MNRAS **468**, 728–740 (2017). 1612.06918.

73. Man, Z.-y. *et al.* The Dependence of AGN Activity on Environment in SDSS. MNRAS 1665 (2019). 1907.01563.

74. Tinker, J. L., Weinberg, D. H., Zheng, Z. & Zehavi, I. On the Mass-to-Light Ratio of Large-Scale Structure. ApJ **631**, 41–58 (2005). arXiv:astro-ph/0411777.

75. White, S. D. M. & Frenk, C. S. Galaxy formation through hierarchical clustering. ApJ **379**, 52–79 (1991).

76. Smith, R. E. *et al.* Stable clustering, the halo model and non-linear cosmological power spectra. MNRAS **341**, 1311–1332 (2003). astro-ph/0207664.

77. Gould, A. chi$\hat{2}$ and Linear Fits. *arXiv e-prints* astro–ph/0310577 (2003). astro-ph/0310577.

78. van Uitert, E., Cacciato, M., Hoekstra, H. & Herbonnet, R. Evolution of the luminosity-to-halo mass relation of LRGs from a combined analysis of SDSS-DR10+RCS2. A&A **579**, A26 (2015). 1503.08647.

79. Tinker, J. L. *et al.* The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests. ApJ **724**, 878–886 (2010). 1001.3162.

80. Klypin, A., Yepes, G., Gottlöber, S., Prada, F. & Heß, S. MultiDark simulations: the story of dark matter halo concentrations and density profiles. MNRAS **457**, 4340–4359 (2016). 1411.4001.

81. Schulze, A. *et al.* The cosmic growth of the active black hole population at 1<z<2 in zCOSMOS, VVDS and SDSS. MNRAS **447**, 2085–2111 (2015). 1412.0754.

82. Lusso, E. *et al.* Bolometric luminosities and Eddington ratios of X-ray selected active galactic nuclei in the XMM-COSMOS survey. MNRAS **425**, 623–640 (2012). 1206.2642.

83. Hopkins, P. F., Richards, G. T. & Hernquist, L. An Observational Determination of the Bolometric Quasar Luminosity Function. ApJ **654**, 731–753 (2007). arXiv:astro-ph/0605678.

84. Zhang, X. & Lu, Y. On the mean radiative efficiency of accreting massive black holes in AGNs and QSOs. *Science China Physics, Mechanics, and Astronomy* **60**, 109511 (2017). 1902.08332.

85. Zhang, X., Lu, Y. & Yu, Q. The Cosmic Evolution of Massive Black Holes and Galaxy Spheroids: Global Constraints at Redshift z < 1.2. ApJ **761**, 5 (2012). 1210.4019.

86. Vasudevan, R. V. & Fabian, A. C. Piecing together the X-ray background: bolometric corrections for active galactic nuclei. MNRAS **381**, 1235–1251 (2007). 0708.4308.

87. Shankar, F., Crocce, M., Miralda-Escudé, J., Fosalba, P. & Weinberg, D. H. On the Radiative Efficiencies, Eddington Ratios, and Duty Cycles of Luminous High-redshift Quasars. ApJ **718**, 231–250 (2010). 0810.4919.

88. Vasudevan, R. V. *et al.* A selection effect boosting the contribution from rapidly spinning black holes to the cosmic X-ray background. MNRAS **458**, 2012–2023 (2016). 1506.01027.

89. Shankar, F., Weinberg, D. H. & Miralda-Escudé, J. Self-Consistent Models of the AGN and Black Hole Populations: Duty Cycles, Accretion Rates, and the Mean Radiative Efficiency. ApJ **690**, 20–41 (2009). 0710.4488.

90. Georgantopoulos, I. & Akylas, A. NuSTAR observations of heavily obscured Swift/BAT AGNs: Constraints on the Compton-thick AGNs fraction. A&A **621**, A28 (2019). 1809.03747.

91. Ananna, T. T. *et al.* The Accretion History of AGNs. I. Supermassive Black Hole Population Synthesis Model. ApJ **871**, 240 (2019). 1810.02298.

92. Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N. & Cen, R. Understanding Black Hole Mass Assembly via Accretion and Mergers at Late Times in Cosmological Simulations. ApJ **799**, 178 (2015).

Acknowledgements FS warmly thanks David Weinberg for a number of illuminating inputs and discussions. FS also acknowledges Dalya Baron, Peter Behroozi, Giorgio Calderone, Benjamin Davis, Fabrizio Fiore, Poshak Gandhi, Sebastian Hoenig, Christian Knigge, Cheng Li, Jordi Miralda-Escudé, Benjamin Moster, Merry Powell, Ranjan Vasudevan, Carolin Villforth, and Guang Yang for many useful discussions. FS acknowledges partial support from a Leverhulme Trust Research Fellowship and the European Union's Horizon 2020 Programme under the AHEAD project (grant agreement n. 654215). VA acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 749348. MB acknowledges partial support from NSF grant AST-1816330. AL is supported by PRIN MIUR 2017 prot. 20173ML3WW_002 "Opening the ALMA window on the cosmic evolution of gas, stars and supermassive black holes". MK acknowledges support from DLR grant 50OR1904.

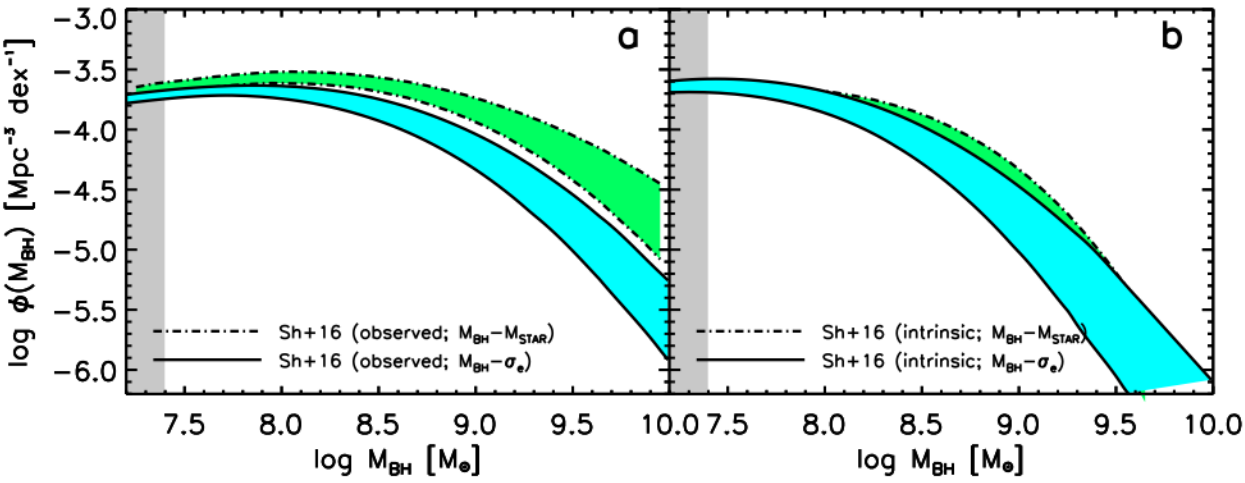

Figure 5: **Predicted local black hole mass functions of elliptical galaxies.** Same format as Figure 3 but only for Elliptical, bulge-dominated galaxies. Lef/right panel show the comparison between the observed/intrinsic $M_{\rm bh}$-$M_{\rm star}$ and $M_{\rm bh}$-σ relations of early-type galaxies, as labelled.

Author contributions statement

All authors have contributed in different ways to the work presented in this paper. FS performed the full set up of the AGN mocks, analysis of the results, and writing up of the manuscript. VA independently checked all the results on AGN clustering and largely contributed to text editing and to the referee reports. MB was one of the core Authors in the Shankar et al. (2016) paper on the intrinsic black hole scaling relations and massively contributed to the editing of the manuscript. CM devised accurate determinations of the correlation functions in the MultiDark simulation. AL calculated the accreted black hole mass functions following the models presented in Aversa et al. (2015). NM contributed to the AGN accretion models. PJG and LZ contributed to the galaxy mocks, independently tested some of the key results in the paper, and provided comments. JM performed preliminary large-scale bias estimates of AGN at different luminosities in low redshift galaxies. MK made available a number of data sets on galaxy and AGN clustering inclusive of full covariance matrices. RDB performed independent calculations of some of the AGN mocks. FR contributed to the characterization of the scaling relations in Type 1 and 2 AGN and to the editing of the paper. FLF provided key insights in the use of different AGN bolometric corrections. RKS was one of the core Authors in the Shankar et al. (2016) paper on the intrinsic black hole scaling relations and provided support on the theoretical side and editing of the paper.

Additional information

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Francesco Shankar (email: F.Shankar@soton.ac.uk).

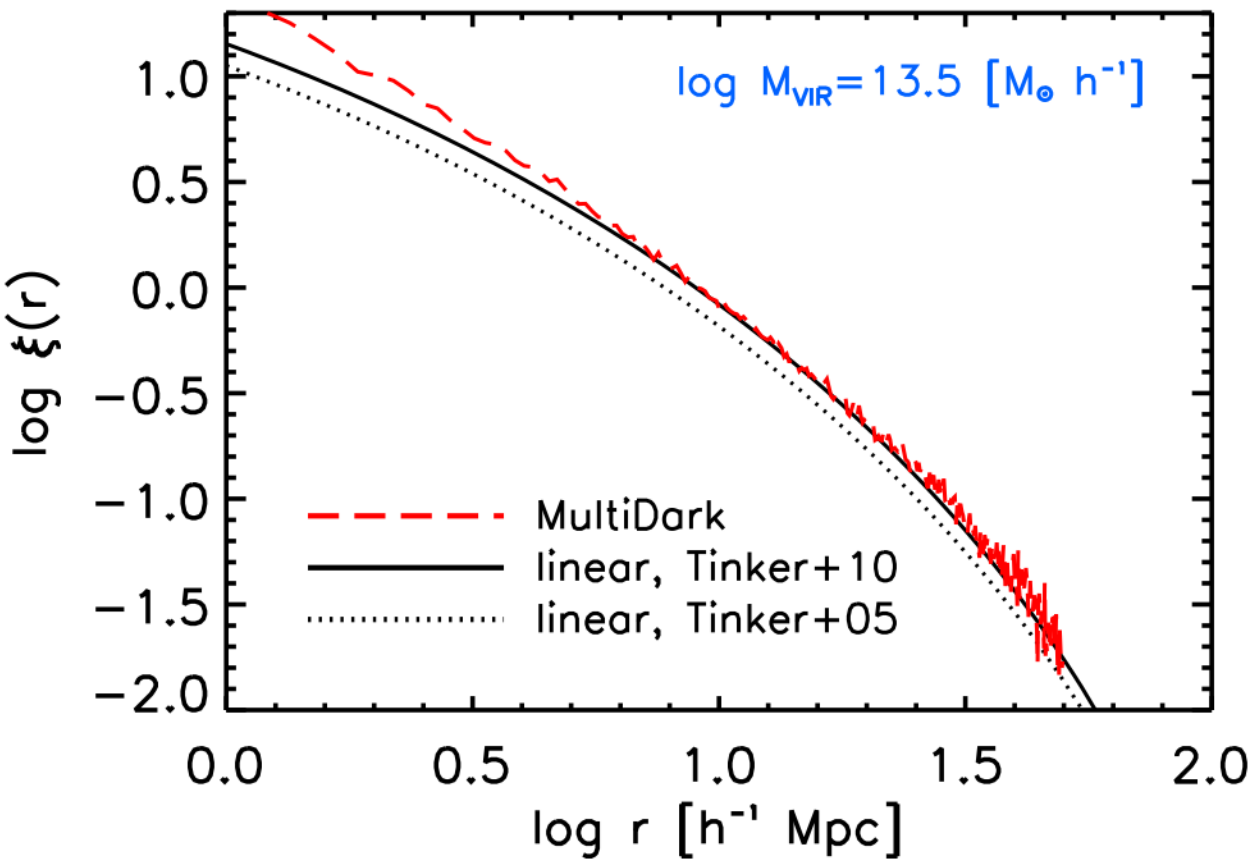

Figure 6: **Comparing host halo auto-correlation functions.** Comparison between the linear matter correlation function multiplied by the square of the halo bias from Tinker et al. (2005; dotted line)[74] and Tinker et al. (2010; solid line)[79], and the autocorrelation function in the MultiDark simulation of all central and satellite haloes with virial mass at infall in the range $13.3 < \log M_{\rm vir}/M_\odot < 13.7$. For this comparison we adopt the same cosmological parameters as in the MultiDark simulation.

$\log M_{\rm bh}$ $[M_\odot]$	$\log \Phi(M_{\rm bh})$ $[h_{70}^{-3}\,{\rm Mpc}^{-3}\,{\rm dex}^{-1}]$
6.0	-2.701 ± 0.015
6.5	-2.730 ± 0.029
7.0	-2.805 ± 0.045
7.5	-2.957 ± 0.073
8.0	-3.218 ± 0.107
8.5	-3.640 ± 0.148
9.0	-4.279 ± 0.193
9.5	-5.233 ± 0.242
10.	-6.678 ± 0.314

Table 1: Black hole mass function retrieved from the intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation.

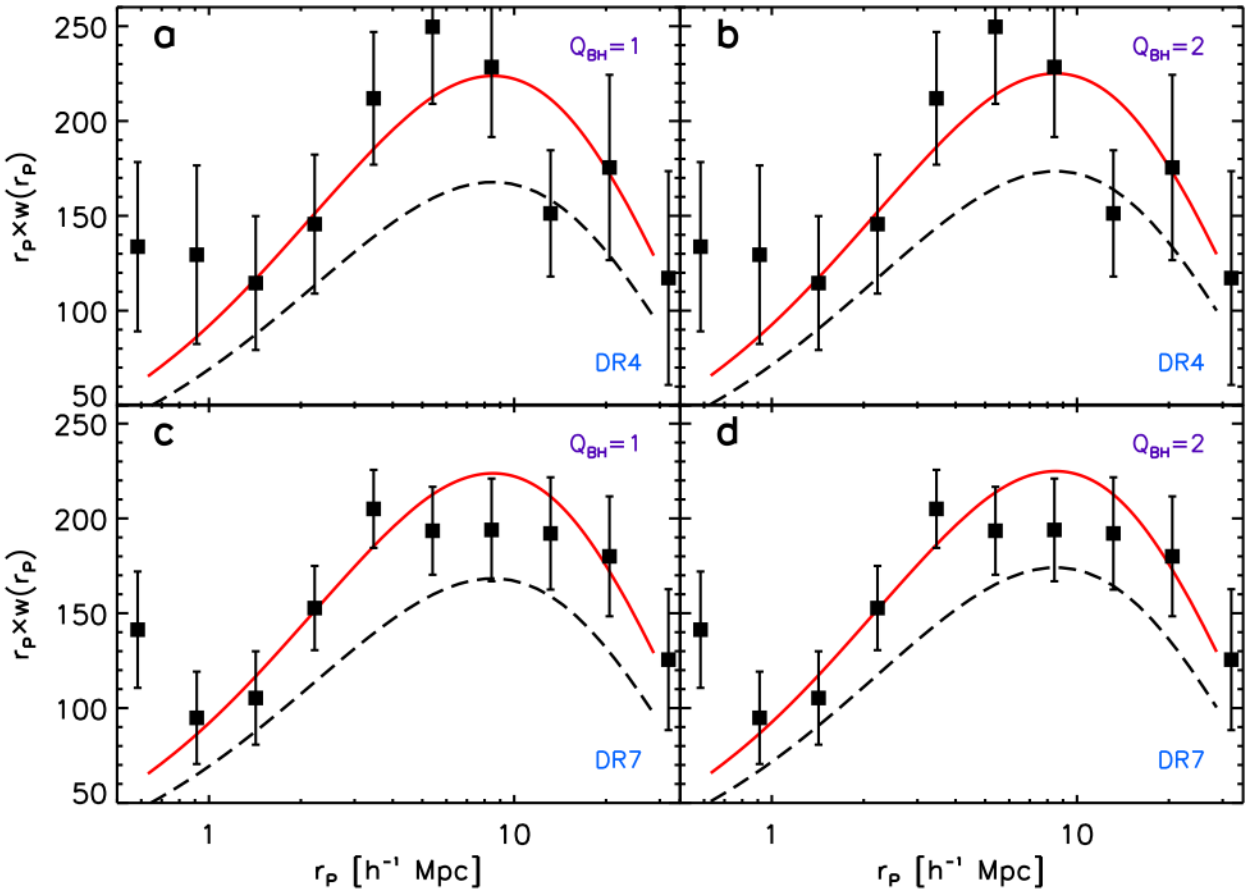

Figure 7: **Direct comparison with the cross-correlation function of active galaxies.** Comparison between the DR4 (top panels) and DR7 (bottom panels) projected AGN-galaxy cross-correlation function derived by Krumpe et al. (filled squares)[24], with the linear matter projected correlation function multiplied by the product of the galaxy and black hole large-scale biases with $Q_{\rm bh} = 1$ (left panel) and $Q_{\rm bh} = 2$ (right panels). It is clear that the models derived from the intrinsic $M_{\rm bh}$-$M_{\rm star}$ relation (solid red lines) provide a better match to the data (see text for details).