The Linear Point Standard Ruler for galaxy survey data: validation with mock catalogues

Stefano Anselmi, Pier-Stefano Corasaniti, Glenn D. Starkman, Ravi K. Sheth, Idit Zehavi

The Linear Point Standard Ruler for galaxy survey data: validation with mock catalogues

Stefano Anselmi*

LUTH, UMR 8102 CNRS, Observatoire de Paris, PSL Research University,
Université Paris Diderot, 92190 Meudon – France and
Institut d'Astrophysique de Paris, CNRS UMR 7095 and UPMC, 98bis, bd Arago, F-75014 Paris – France

Pier-Stefano Corasaniti

LUTH, UMR 8102 CNRS, Observatoire de Paris, PSL Research University,
Université Paris Diderot, 92190 Meudon – France

Glenn D. Starkman

Department of Physics/CERCA/Institute for the Science of Origins,
Case Western Reserve University, Cleveland, OH 44106-7079 – USA

Ravi K. Sheth

Center for Particle Cosmology, University of Pennsylvania,
209 S. 33rd St., Philadelphia, PA 19104 – USA and
The Abdus Salam International Center for Theoretical Physics, Strada Costiera, 11, Trieste 34151 – Italy

Idit Zehavi

Department of Astronomy, Case Western Reserve University, Cleveland, OH 44106-7079 – USA

(Dated: January 16, 2018)

Due to late time non-linearities, the location of the acoustic peak in the two-point galaxy correlation function is a redshift-dependent quantity, thus it cannot be simply employed as a cosmological standard ruler. This has motivated the recent proposal of a novel ruler, also located in the Baryon Acoustic Oscillation range of scales of the correlation function, dubbed the *linear point*. Unlike the peak, it is insensitive at the 0.5% level to many of the non-linear effects that distort the clustering correlation function and shift the peak. However, this is not enough to make the linear point a useful standard ruler. In addition, we require a model-independent method to estimate its value from real data, avoiding the need to deploy a poorly known non-linear model of the correlation function. In this manuscript, we precisely validate a procedure for model-independent estimation of the linear point. We also identify the optimal set-up to estimate the linear point from the correlation function using galaxy catalogs. The methodology developed here is of general validity, and can be applied to any galaxy correlation-function data. As a working example, we apply this procedure to the LOWZ and CMASS galaxy samples of the Twelfth Data Release (DR12) of the Baryon Oscillation Spectroscopic Survey (BOSS), for which the estimates of cosmic distances using the linear point have been presented in Anselmi et al. (2017) [1].

I. INTRODUCTION

Baryon Acoustic Oscillations (BAO) in the late-time matter power spectrum result from primeval acoustic waves propagating in the coupled baryon-photon plasma before decoupling [2–4]. These manifest as a peak in the two-point correlation function (CF) of galaxies, located at the scale of the sound horizon at the so-called drag epoch, when the acoustic waves stop freely propagating through the plasma. This provides a natural comoving standard ruler to constrain the cosmic expansion history [5–7].

Ideally, one would estimate cosmic distances by measuring the location of the BAO peak directly from CF data, without the need to model the processes that shape the CF. Unfortunately, on BAO scales the late-time distribution of matter is sensitive to the non-linear dynamics of matter's gravitational clustering. Several studies, using both high-precision cosmological simulations and analytic models, have shown that non-linearities distort the BAO pattern: smearing the BAO peak, lowering its amplitude and shifting its position [8–11]. Therefore peak-finding algorithms cannot be just blindly applied to the data to extract cosmic distance information, rather the opposite – one should use cosmology-dependent fits of the full CF [12, 13]. This would be a minor inconvenience if we knew how to predict the full non-linear galaxy CF as a function of only the cosmological parameters. Unfortunately, we are far from achieving that goal.

In the past ten years, several approximate methods have been developed to extract cosmic distance information from BAO measurements. The most widely accepted technique defines the BAO scale in terms of a

*Electronic address: stefano.anselmi@iap.fr

fiducial-model template CF, where the cosmological parameters are kept fixed at the fiducial values. *Ad hoc* nuisance parameters are added, to capture the effects of non-linearities and with the intent of "marginalizing over" the chosen fiducial cosmology [14, 15]. This model template is then used to infer the cosmic distance from the statistical data analysis. Moreover, since non-linearities suppress the amplitude of the BAO, the observed galaxy positions are adjusted, using approximate non-linear model algorithms, to enhance the signal-to-noise of the BAO peak in the CF. This is done with the intent of restoring the pristine information on the acoustic scale; however, this *reconstruction* procedure explicitly depends on the choice of a fiducial cosmology and on the specification of a heuristic model of non-linear effects [16]. Hence, in both the treatment of the data and the statistical analysis, model-dependent assumptions intervene. These carry the inherent risk of underestimating the uncertainties on cosmic distances, and potentially introduce a source of systematic bias in the cosmological-parameter inference.

In order to overcome these limitations, a new promising BAO standard ruler in the galaxy CF, dubbed the *linear point* (LP hereafter), was suggested by some of us [17]. Its position, defined as the midpoint between the positions of the peak and dip in the monopole CF, is located at ~ 95 Mpc/h in comoving units [17]. Using results from N-body simulations of ΛCDM models, it has been shown that the LP is insensitive to non-linear effects at 0.5% relative to the linear-theory prediction. This holds for the matter-density field as well as for the spatial distribution of halos. Moreover, analytic arguments suggest that the LP remains stable (in both position and amplitude) with respect to the effects of redshift-space distortions and scale-dependent bias [17]. An additional advantage of the LP is that it is a purely geometrical standard ruler, i.e., its position is independent of the amplitude and slope of the spectrum of primordial density fluctuations (at least for models similar to the ΛCDM scenario). Hence, unlike any other known BAO analysis, the LP can provide estimates of cosmic distances without the need for theoretical modelling of the CF data.

Recently, we have presented [1] a cosmological relation that allows us to infer the isotropic-volume distance D_V using estimates of the LP from galaxy data. In particular, we focused on the CMASS and LOWZ galaxy samples from the Twelfth Data Release (DR12) of the Baryon Oscillation Spectroscopic Survey (BOSS)[1], and found

$$D_V^{\rm LP}(\bar{z}_{\rm LOWZ-DR12} = 0.32) = (1264 \pm 28)\,{\rm Mpc}\,,$$
$$D_V^{\rm LP}(\bar{z}_{\rm CMASS-DR12} = 0.57) = (2056 \pm 22)\,{\rm Mpc}\,, \quad (1)$$

thus providing distance estimates that are competitive with those obtained from standard assumption-rich BAO methods.

In this manuscript, we aim to validate the LP parametric model-independent estimation already applied in [1] to the actual LOWZ and CMASS galaxy samples. To this end, we employ the Quick Particle Mesh (QPM) mock catalogues ("mocks") [18] built by the BOSS collaboration explicitly to mimic the LOWZ and CMASS clustering properties. They were largely used by the collaboration to test their Twelfth Data Release (DR12) BAO data analysis [19].

Our approach relies on a simple polynomial interpolation of the CF in the BAO range of scale. In this paper, we first validate the polynomial fit. Then, for each mock, the best-fit polynomial parameters and uncertainties provide the LP estimate and error. We find the optimal values of the polynomial order, the fitted range of scales, and the bin size to use for LP estimation on this data set. Optimization for future, larger-volume or higher-precision data sets would yield different values; but remarkably, our preliminary tests suggest that it will be sufficient just to shrink the fitted range of scales.

The paper is structured as follows. In Section II, we detail the methodology employed to validate the linear-point estimation through the polynomial fit: we summarize the characteristics of the QPM mocks, we define the systematic bias, we provide a checklist that the optimal fitting set-up should pass to be validated. In Section III, we perform the previously introduced tests, discussing step by step the results of the analysis. In Section IV, we present our conclusions.

II.   METHODOLOGY

In this section, we present the procedure developed to estimate the linear point from galaxy data. Our goal is to show that a simple model-independent parametric fit, applied to the monopole clustering correlation function, recovers the LP position without introducing systematic biases. We test this on mock catalogues, generated by the BOSS collaboration to reproduce the Luminous Red Galaxies (LRG) DR12-BAO clustering properties and used to test the BOSS BAO analysis [18, 19][2]

A.   QPM Mocks

QPM mocks [18] were employed for the BOSS clustering analysis. The QPM method uses a low-resolution particle-mesh N-body solver. The halo catalogue and its properties were built to match the mass function and

---

large-scale bias of halos of high-resolution simulations. The halo catalogue was then populated with galaxies using a halo occupation distribution (HOD) modelling, where the HOD parameters were adjusted for each mock by fitting the observed small-scale projected two-point galaxy correlation function for the LRGs. Each mock matches the angular and radial selection functions of the survey and the observed number density of galaxies. The final galaxy catalogue consists of 1000 realizations of the LOWZ sample and 956 for CMASS [17]. For each of these mocks, the CF has been computed using the Landy-Szalay algorithm [21].

The fiducial cosmology of the QPM mocks is a flat ΛCDM model, with cosmological parameter values close to the best-fit Planck+BOSS cosmology: $\Omega_m = 0.29$, $\Omega_\Lambda = 0.71$, $\Omega_b h^2 = 0.02247$, $\Omega_\nu h^2 = 0.0$, $h = 0.7$, $n_s = 0.97$, and $\sigma_8 = 0.8$.

B. Estimating the Linear Point position with a model-independent parametric fit

In order to extract the LP position from the galaxy monopole correlation function $\xi_0(s)$ (s being the redshift-space coordinate in comoving units), we first estimate the positions of the maximum and the minimum of the CF in the BAO range of scales. This can readily be accomplished using a model-independent parametric fit. A simple, but (as we will see) efficient and robust, way to do so consists of first interpolating the CF data with a polynomial

$$\xi_0^{\text{fit}}(s) = \sum_{i=0}^{n} a_i s^i, \tag{2}$$

where n is the order of the polynomial fitting function. The solutions of $d\xi_0^{\text{fit}}/ds = 0$ are then computed, to find the location of the peak ($\hat{s}_{\text{peak}}^{\text{fit}}$) and dip ($\hat{s}_{\text{dip}}^{\text{fit}}$) in the CF.[3] The estimated location of the LP is the mid-point between the computed dip and peak locations

$$\hat{s}_{LP}^{\text{fit}} = \frac{1}{2}(\hat{s}_{\text{peak}}^{\text{fit}} + \hat{s}_{\text{dip}}^{\text{fit}}), \tag{3}$$

which can be expressed in terms of the best-fit polynomial coefficients to the CF data.[4] This allows us to estimate the uncertainty on the LP location, by propagating the uncertainties in the polynomial coefficients.

––––––––

[3] It is worth noting that, in the analysis of real (rather than simulated) galaxy-survey data, one should account for the Alcock-Paczynski effect [22, 23], which distorts the CF. In such a case, one can conveniently express the CF in terms of the dimensionless distance $y \equiv s/constant$ [13, 17]. However, the procedure to extract the LP is the same whether the correlation function is expressed as function of y or s. Therefore, to ease the reading of the present article, we work in comoving coordinates.

[4] To simplify the notation hereafter, we omit the hat and the "fit" subscripts.

We would like to stress two considerations concerning the use of the polynomial interpolation of the CF. First, it provides an effective way of smoothing the noisy data points, thereby enabling the LP estimation. Indeed the more parameters we allow (i.e., the higher the order of the polynomial) the less effectively the polynomial fit smooths the CF. Nevertheless, we expect that the fitting procedure does not introduce a systematic bias in the determination of the LP, as we will show in Section III B. Second, the authors of [17] found that, in the BAO range of scales, the CF is nearly anti-symmetric with respect to the LP. As we will show in Section III C, this provides us with a guideline to choose the optimal range of scales over which to interpolate the CF.

In principle, the order of the CF polynomial-fitting function may depend on the range of scales considered, the redshift and the survey volume. Here, we find that an unbiased estimator of the LP requires $n \geq 5$. In the case of the LOWZ and CMASS mocks, we find that a quintic polynomial fit the CFs well over the range of scales considered. We will show this in Section III B, by comparing to the LP estimate obtained using a seventh-order polynomial.

C. Linear Point estimation: bias definition

Our analysis has two goals: on the one hand, we want to show that a simple polynomial fit to the CF can provide an unbiased estimate of the LP; on the other hand, we want to determine the optimal combination of polynomial order, range of scales, and binning that minimizes the LP statistical error. To this end, we introduce a measure of the LP systematic bias:

$$b_{LP} = \bar{s}_{LP} - s_{LP}^{\text{true}}. \tag{4}$$

Here, $\bar{s}_{LP}$ is the mean of the LP positions estimated from the mocks, and s_{LP}^{true} is our reference LP value, which we set to the value of the LP estimated from the average CF over the mocks. This is because we are interested in evaluating only the uncertainty in the LP estimation due to the polynomial-interpolation procedure, and not any small uncertainty that arises due to the non-linear clustering of matter [17]. As already mentioned in Section I, that non-linear clustering systematically shifts the location of the LP relative to the linear CF up to a 0.5% [17], but this can be mitigated by shifting the definition of the LP estimator relative to Eq. (3). (This shift leaves Eq. (4) unaffected, since it affects $\bar{s}_{LP}$ and s_{LP}^{true} equally.) We therefore take s_{LP}^{true} to be the value estimated from the CF averaged over all the mocks. This has negligible cosmic-variance and sampling-variance error compared to the individual mocks, but (deliberately)
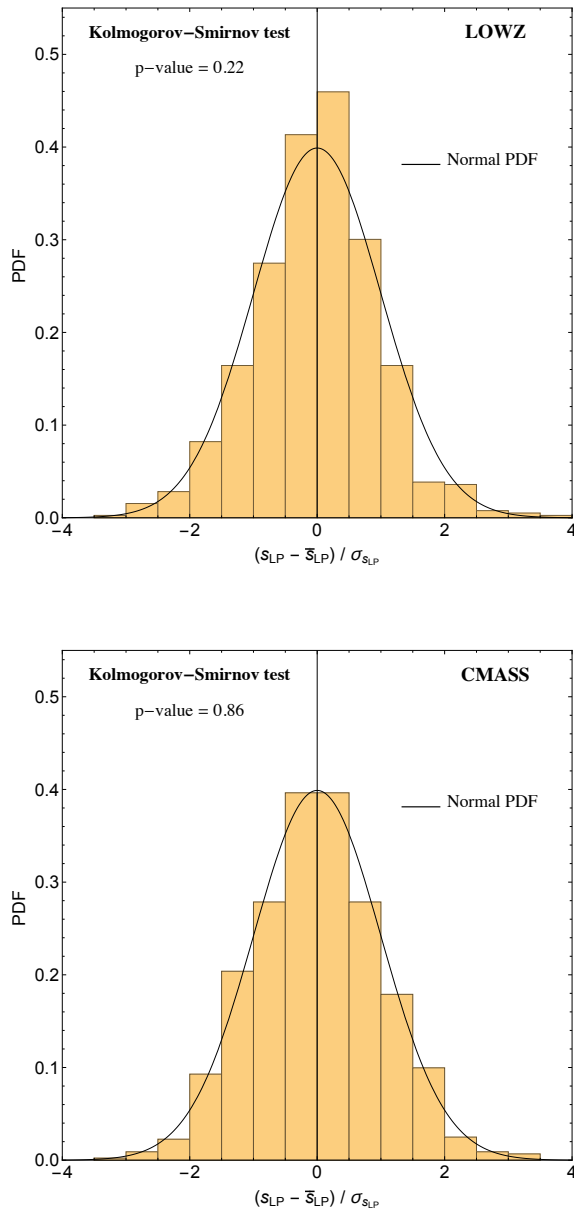
FIG. 1: Normalized histograms of the rescaled linear-point positions recovered from the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogues. The unit normal probability distribution function is overplotted. The *p*-values of the Kolmogorov-Smirnov test show that, for both catalogues, there is a reasonable probability that the LP values are drawn from a Gaussian distribution.

shares with them any potential systematic bias in $\bar{s}_{LP}$.[5]

———————

[5] In [17], we have shown that the LP position in the CF, for both high-resolution N-body simulations and theoretical models, shifts with respect to the linear theory prediction by no more than 1%. This shift is secular, and its effect is halved with a simple redshift-independent correction. In the case of the LOWZ and CMASS mocks, we find the "true" LP position to deviate by

We would like to highlight that, given the finite volume of the mocks, the BAO feature in the CF might not be present in a given mock, due to cosmic variance. In this case, the polynomial estimator would not "detect" the peak and dip in the BAO range of scale, i.e., $d\xi_0^{\rm fit}/ds = 0$ would have no actual solutions. Since the BAO is detected in the real data, we condition the LP error analysis to the mocks in which the LP is detected (for each polynomial-fit configuration). We therefore compute the conditional CF data covariance recursively: we perform a first polynomial fit of the CF of each mock using the covariance from the entire mock dataset, and, if the LP is not detected, we discard the mock and recompute the CF data covariance from the selected mocks. Depending on the polynomial fit configuration (order of the polynomial, bin size and range of scales), the fraction of retained mocks (Mock Acceptance Rate) is $\gtrsim 80\%$ for LOWZ and $\gtrsim 90\%$ for CMASS. As the statistical error on the LP position and bias are normalized to the number of retained mocks, the LP error analysis is independent of the specific value of the mock acceptance rate. Moreover, since there is no LP-position information in the mocks that do not show the BAO features, including them in the covariance-matrix computation does not change the LP error estimate.

Given the finite size of the mock samples, we correct the LP error budget according to [26, 27]. Similarly, we follow [27] for estimating errors on the determination of the fitting-polynomial coefficients for each mock's CF.

In summary, our validation of the LP estimation will assess the following points:

A. Gaussianity of the correlation-function and linear point distributions: we show that the both the CF and the LP are consistent with a Gaussian distribution.

B. Optimal polynomial estimator: we consider polynomials of different orders as LP estimators, and discuss their suitability.

C. Optimal BAO range of scales: we analyze the BAO range of scale fit for the CF, and identify the optimal one for LP estimation.

———————

1.3% and 1.2% w.r.t. the linear theory prediction respectively. This disagreement could be due to the approximate treatment of clustering in the modeling used to built the QPM mocks [18]. Alternately, it might be due to the way the HOD model is implemented in the QPM mocks. For instance, rather than adjusting the HOD parameters separately for each mock, one should properly fit the projected CF on small scales once for all the mocks (see e.g. [24]). Furthermore, the parameter uncertainties due to the HOD fitting should be correctly propagated to the BAO scales. We plan to properly address these points with further investigations to be carried out with high-resolution and large-volume N-body simulations [25] and improved implementations of the HOD [24]. Nevertheless, for the purpose of the present analysis, which is validating the non-linear LP extraction, we can safely ignore this issue.

Estimator Test

Polyn.	b_{LP}	$\bar{\sigma}_{s_{LP}}$	mean AIC$_c$
- LOWZ			
Quintic	-0.41 Mpc/h	2.4 Mpc/h	34
7th order	-0.37 Mpc/h	2.7 Mpc/h	41
- CMASS			
Quintic	-0.25 Mpc/h	1.5 Mpc/h	35
7th order	-0.20 Mpc/h	1.7 Mpc/h	41

TABLE I: We show the results of the *estimator test*. Both the quintic and the seventh-order polynomials are unbiased (i.e., negligible-bias) linear-point estimators. The quintic polynomial is the chosen LP estimator, as it provides the smallest errors and is preferred by the model-selection criterion.

D. Optimal bin size: we identify the bin sizes that return an unbiased LP estimate.

III. LINEAR POINT ESTIMATION TESTS

As mentioned in Section I, the advantage of the LP is that it is a geometric standard ruler on the BAO scale that is preserved by non-linear effects. LP estimation therefore does not require the use of reconstruction methods to be applied to galaxy catalogs. Hence, we test the LP estimation procedure on pre-reconstructed QPM mocks from the BOSS collaboration.

The results of the error evaluation, which will be presented below, indicate that the optimal setup for LP estimation consists of fitting the galaxy CF with a quintic polynomial estimator, in the range of scales $60 < s$ [Mpc/h] < 130, with bins of size $\Delta s = 3$ Mpc/h.

A. Gaussianity of the correlation function and linear point distributions

We verify that the distribution of the mock CF is always well-described by a Gaussian function. Hence, we can perform a statistical analysis of the CF assuming a Gaussian likelihood, and find the coefficients of a polynomial fitting function by simple χ^2 minimization.

We check that the distribution of the $\chi^2_{\min}$ values from the polynomial fit to the CF of the mocks is consistent with a χ^2 distribution, while the distribution of inferred values of s_{LP} is consistent with a Gaussian. The latter is shown in Fig. 1, where we plot the LOWZ *(upper panel)* and CMASS *(bottom panel)* normalized histograms of $(s_{LP} - \bar{s}_{LP})/\sigma_{s_{LP}}$. The unit normal probability distribution function is overplotted. We perform
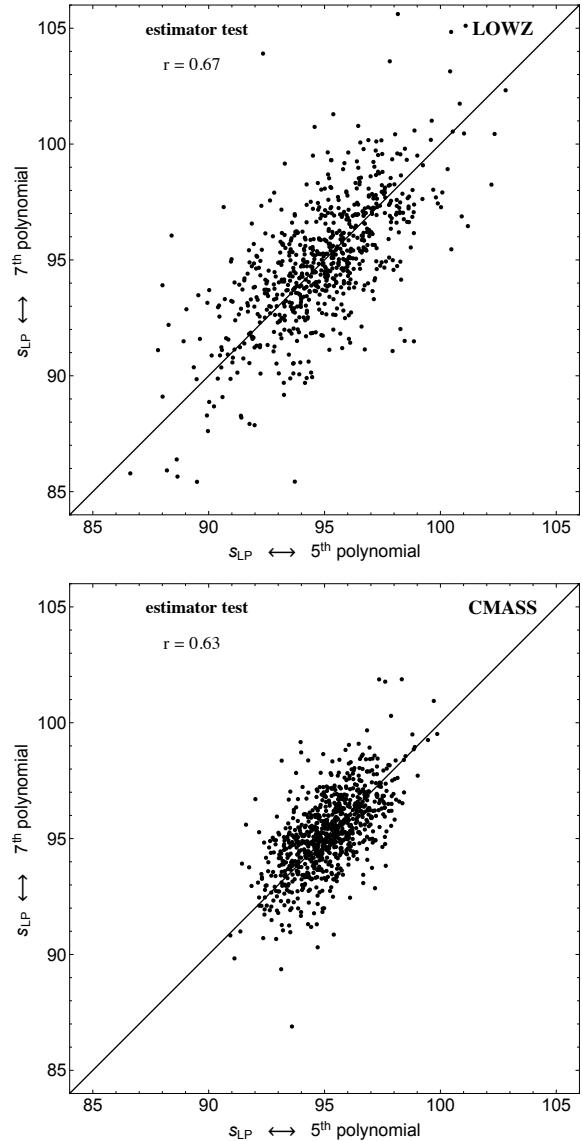

FIG. 2: Scatter in the LP *estimator* for the quintic *versus* the seventh-order polynomial fits to the CF of the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogues. The scatter along the continuous black line indicates the cosmic-variance error, while the scatter perpendicular to the line represents the estimator error. The larger values of the correlation coefficients quoted in the panels suggest that the errors on the LP estimator are dominated by cosmic variance.

the Kolmogorov-Smirnov test for the LOWZ and CMASS mocks. The p-values are respectively 0.22 and 0.86, indicating reasonable probabilities that the posterior of s_{LP} is Gaussian distributed. Therefore, we can assign the usual Gaussian meaning to the rms of the LP distribu-

tion.[6]

The mean s_{LP} errors for the two simulated galaxy samples are $\sigma_{s_{LP}}^{\mathrm{LOWZ}} = 2.4$ Mpc/h and $\sigma_{s_{LP}}^{\mathrm{CMASS}} = 1.5$ Mpc/h. Therefore, the intrinsic 0.5% deviation of s_{LP} with respect to s_{LP}^{lin} (found in [17]) is subdominant.
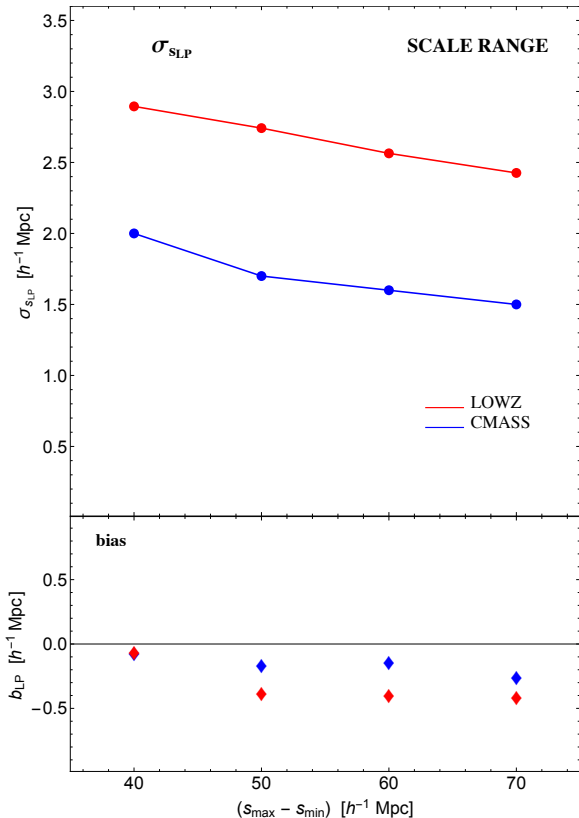

FIG. 3: LP-estimation error *(upper panel)* and bias *(lower panel)* as functions of the range of scales used to fit the CF. As we can see, the bias is always negligible, since $b_{LP} \leq 0.2 \times \bar{\sigma}_{s_{LP}}$.

B. Optimal polynomial estimator

To test the dependence of the LP estimation on the choice of the order of the polynomial fit to the CF, we consider a quintic polynomial and a seventh-order one. For each of these cases, we estimate s_{LP} from Eq. (3) and evaluate the bias as in Eq. (4). The results of the comparison are summarized in Table I, where we quote, for each mock, both the value of the bias and the average error on the LP estimator. In all cases, the absolute

---

systematic shift is much smaller than the mean error estimated from the likelihood, i.e., $b_{LP} < 0.2 \times \bar{\sigma}_{LP}$. As expected, $\bar{\sigma}_{s_{LP}^{5\mathrm{th}}} < \bar{\sigma}_{s_{LP}^{7\mathrm{th}}}$, since the quintic-polynomial interpolation needs to fit a smaller number of parameters.

In Fig. 2, we show the scatter-plots of the recovered LP position for the two polynomial orders, using the LOWZ *(upper panel)* and CMASS *(lower panel)* mocks. The scatter along the solid line is an indication of the cosmic-variance error, while the scatter perpendicular to the solid line indicates the estimator error. Quantitatively, Pearson's correlation coefficient r for the two samples, $r^{\mathrm{LOWZ}} = 0.67$ and $r^{\mathrm{CMASS}} = 0.63$, reveals that cosmic variance is the dominant error. We have checked that combined use of the two estimators does not significantly reduce the statistical error compared with using only the quintic-polynomial fit. Therefore, using only the quintic polynomial is sufficient for our purposes.

Since both of the estimators are unbiased, to choose between them we adopt a simple model-selection criterion: the finite-sample-corrected Akaike information criterion $\mathrm{AIC_c}$ [28]. We report its formula here for convenience (dropping an irrelevant additive constant):

$$\mathrm{AIC_c} \equiv \chi_{\min}^2 + \frac{2(n+1)N}{N-n-2}, \qquad (5)$$

where n was introduced in Eq. (2) and N is the number of points fit. The idea behind the $\mathrm{AIC_c}$ is to balance the quality of fit to the observed data against the complexity of the model. The polynomial fit that gives the minimal $\mathrm{AIC_c}$ value is selected. From Table I, we see that the smallest mean $\mathrm{AIC_c}$ belongs always to the quintic polynomial. This motivates its choice for the optimal set-up.

C. Optimal BAO range of scales

We focus next on determining the optimal range of scales from which to extract the LP from measurements of the CF.

The values of the coefficients of the best-fit polynomial to the CF, and their associated errors, depend on the range of scale over which the CF is fit. This calls for selecting an optimal range of scales that minimizes the statistical uncertainty, while introducing negligible systematic bias in the estimated LP location. We recall that the CF is anti-symmetric with respect to the linear point over the BAO range of scales [17]. This motivates interpolating the CF with a quintic polynomial over a range that is symmetric with respect to 95 Mpc/h.

In Fig. 3, we plot the statistical error *(upper panel)* and bias *(lower panel)* in the LP position, as functions of the interval of scales over which the CF is interpolated. We observe that, when the CF is interpolated over $(s_{\max} - s_{\min}) = 70$ Mpc/h, the statistical error in the LP is minimized, while the systematic bias is negligible, $b_{LP} \leq 0.2 \times \bar{\sigma}_{s_{LP}}$. This trend is expected, as the fitting parameters are better determined when more information from the data is included.

---

[6] We recall that, in this manuscript, we always use the error estimated from the likelihood and not from the distribution. However, after applying the corrections for the small number of mocks [27], the two agree to better than 7%.
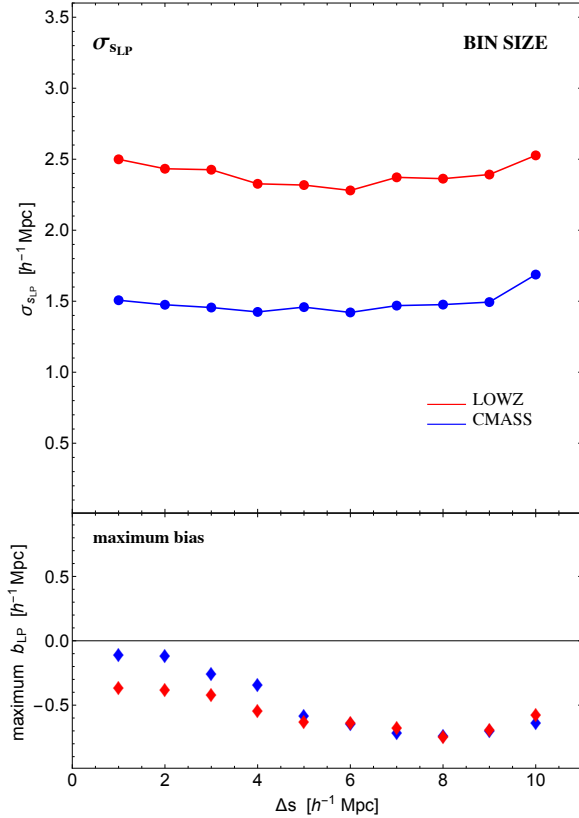
FIG. 4: LP-estimation error *(upper panel)* and bias *(lower panel)* as functions of the bin width. We see that the bias in the LP estimate induced by the choice of the binning of the CF is negligible (i.e., $b_{LP} \leq 0.2 \times \bar{\sigma}_{s_{LP}}$) for $\Delta s \leq 4$ Mpc/h.

We do not explore wider ranges of scale, since for the LOWZ real galaxy data, due to the low signal-to-noise, the LP is detected only for $(s_{\max} - s_{\min}) \leq 70$ Mpc/h [1]. For the CMASS mock data, extending the fit over a larger range of scales does not result in a further reduction of the statistical errors. Thus we conclude that the optimal range of scales to is $(s_{\max} - s_{\min}) = 70$ Mpc/h.

D. Optimal bin size

The CF is measured in bins of finite width from datasets, whether simulated or real. In principle, the binning procedure can affect the LP estimation. To assess this effect, we have considered bins of varying size, from $\Delta s = 1$ Mpc/h to $\Delta s = 10$ Mpc/h, and rebinned the CF mock data accordingly. In Fig. 4, we plot the values of $\sigma_{s_{LP}}$ and b_{LP} corresponding to the most biased result among all the possible CF sampling possibilities for that bin size. While the LP statistical uncertainty is largely independent of the bin size, for $\Delta s \geq 5$ Mpc/h the mean LP value recovered from the mocks can be significantly biased. A too-large bin size introduces uncertainties in the bin positions, it does not provide enough sampling of
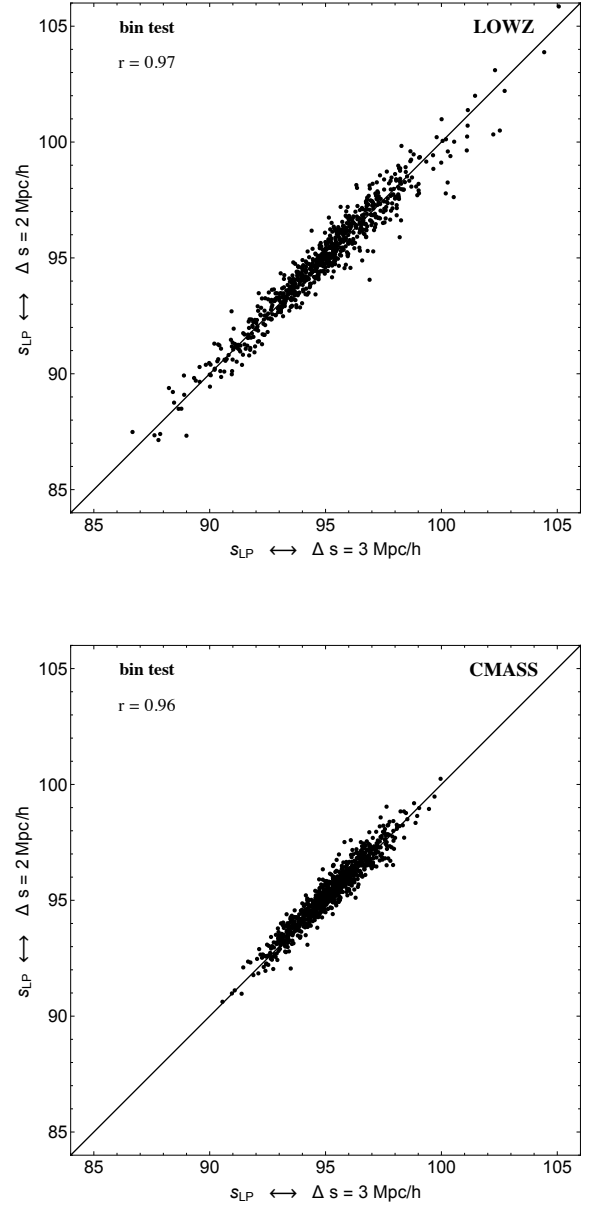

FIG. 5: LP-estimator scatter plot, for a quintic-polynomial fit to the CF, with bins of width $\Delta s = 2$ and 3 Mpc/h, for the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogues.

the CF in the BAO range of scales, and it introduces a dependence on the sampling choice.

In Fig. 5, we show that the recovered LP position exhibits small scatter for small bin-sizes and, consequently, a high correlation coefficient r between the $\Delta s = 2$ Mpc/h and the $\Delta s = 3$ Mpc/h LP estimators.

We conclude that, for $\Delta s \leq 4$ Mpc/h, the LP systematic bias is negligible; hence, recalling that a larger bin-size allows to reduce the covariance matrix noise [27], we choose $\Delta s = 3$ Mpc/h for the optimal set-up.

IV. CONCLUSIONS

Equipping the Baryon Acoustic Oscillations with a cosmological standard ruler is a highly desirable goal. It must be independent of the parameters characterizing the primordial fluctuations (within inflationary ΛCDM), and insensitive to non-linearities that develop during the late-time dark-energy-dominated era. The *linear point* provides such a ruler.

Another feature of the LP was not previously considered: its simple definition allows a model-independent BAO analysis. In this work, using mock galaxy catalogs, we have presented a validation of LP-estimation through a theory-free parametric fit to the galaxy CF. In [1], we applied such an estimator to galaxy data, and showed that the method presented here holds, even when the Alcock-Paczynski distortion is present. We thus discovered that cosmological distances can be estimated without any need to model the non-linear physics that affects the galaxy correlation function at the BAO range of scales.

In this paper, we have determined the optimal set-up to extract, by means of the LP, distance information from the BOSS-DR12 LOWZ and CMASS galaxy samples, justifying the methodology applied in [1]. This consists in using a quintic polynomial to fit the galaxy CF over the range of scales $60 < s < 130$ Mpc/h, with a bin width of $\Delta s = 3$ Mpc/h.

We have noted that the LP observable is not detected in all the available mocks, because, for a small fraction of the mocks, the polynomial fits do not exhibit the peak-and-dip feature necessary for polynomial LP estimation. Since fits to both the LOWZ and CMASS data do exhibit the needed peak and dip [1], we consistently condition our analysis to the mocks compatible with observations. It is important to stress that, as can be easily tested, this procedure does not change the LP error budget, since the excluded mocks do not carry information on the LP position.

We plan to perform a LP-standard-ruler forecast analysis for future galaxy surveys such as Euclid (`http://sci.esa.int/euclid/`), DESI (`http://desi.lbl.gov`) and WFIRST (`https://wfirst.gsfc.nasa.gov`). The LP is also promising as a probe of the growth of structure, given that the amplitude of the CF at the LP is insensitive to non-linearities [17]. Also worth investigating is the effect on the LP of massive neutrinos [29, 30], which is still not considered even in standard BAO analysis. The LP may also serve as a smoking gun of modified-gravity, especially if the BAO-LP anti-symmetric feature [17] is altered in candidate models (such as Quasidilaton Massive Gravity Theory [31, 32]). Alternatively, one could construct a maximal-deviation test where, in the context of the concordance ΛCDM, the maximal allowed deviation from the predicted CF anti-symmetry feature (in the BAO regime) is quantified.

Acknowledgments

[1] S. Anselmi, G. D. Starkman, P.-S. Corasaniti, R. K. Sheth, and I. Zehavi, ArXiv e-prints (2017), 1703.01275.

[2] P. J. E. Peebles and J. T. Yu, Astrophys. J. **162**, 815 (1970).

[3] A. D. Sakharov, Soviet Journal of Experimental and Theoretical Physics **22**, 241 (1966).

[4] J. Silk, Astrophys. J. **151**, 459 (1968).

[5] D. J. Eisenstein, W. Hu, and M. Tegmark, Astrophys. J. Lett. **504**, L57 (1998), astro-ph/9805239.

[6] B. A. Bassett and R. Hlozek (2009), 0910.5224.

[7] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, et al., Astrophys. J. **633**, 560 (2005), astro-ph/0501171.

[8] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **77**, 043525 (2008), astro-ph/0703620.

[9] D. J. Eisenstein, H.-J. Seo, and M. White, Astrophys. J. **664**, 660 (2007), astro-ph/0604361.

[10] T. Matsubara, Phys.Rev. **D77**, 063530 (2008), 0711.2521.

[11] M. Crocce and R. Scoccimarro, Phys. Rev. **D77**, 023533 (2008), 0704.2783.

[12] A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga, Mon. Not. R. Astron. Soc. **400**, 1643 (2009), 0901.2570.

[13] A. G. Sanchez, C. Scoccola, A. Ross, W. Percival, M. Manera, et al., Mon.Not.Roy.Astron.Soc. **425**, 415 (2012), 1203.6616.

[14] H.-J. Seo, E. R. Siegel, D. J. Eisenstein, and M. White, Astrophys. J. **686**, 13 (2008), 0805.0117.

[15] X. Xu, N. Padmanabhan, D. J. Eisenstein, K. T. Mehta, and A. J. Cuesta, Mon. Not. R. Astron. Soc. **427**, 2146 (2012), 1202.0091.

[16] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, Astrophys. J. **664**, 675 (2007), astro-ph/0604362.

[17] S. Anselmi, G. D. Starkman, and R. K. Sheth, Mon. Not. R. Astron. Soc. **455**, 2474 (2016), 1508.01170.

[18] M. White, J. L. Tinker, and C. K. McBride, Mon. Not. R. Astron. Soc. **437**, 2594 (2014), 1309.5532.

[19] A. J. Cuesta, M. Vargas-Magaña, F. Beutler, A. S. Bolton, J. R. Brownstein, D. J. Eisenstein, H. Gil-Marín, S. Ho, C. K. McBride, C. Maraston, et al., Mon. Not. R.

Astron. Soc. **457**, 1770 (2016), 1509.06371.

[20] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, et al., Mon. Not. R. Astron. Soc. **470**, 2617 (2017), 1607.03155.

[21] S. D. Landy and A. S. Szalay, Astrophys. J. **412**, 64 (1993).

[22] C. Alcock and B. Paczynski, Nature (London) **281**, 358 (1979).

[23] X. Xu, A. J. Cuesta, N. Padmanabhan, D. J. Eisenstein, and C. K. McBride, Mon. Not. R. Astron. Soc. **431**, 2834 (2013), 1206.6732.

[24] M. Sinha, A. A. Berlind, C. K. McBride, R. Scoccimarro, J. A. Piscionere, and B. D. Wibking, ArXiv e-prints (2017), 1708.04892.

[25] Y. Rasera, P.-S. Corasaniti, J.-M. Alimi, V. Bouillot, V. Reverdy, and I. Balmès, Mon. Not. R. Astron. Soc. **440**, 1420 (2014), 1311.5662.

[26] J. Hartlap, P. Simon, and P. Schneider, A&A **464**, 399 (2007), astro-ph/0608064.

[27] W. J. Percival, A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, R. Crittenden, A. J. Cuesta, M. V. Magana, M. Manera, F. Beutler, et al., Mon. Not. R. Astron. Soc. **439**, 2531 (2014), 1312.4841.

[28] A. R. Liddle, Mon. Not. R. Astron. Soc. **377**, L74 (2007), astro-ph/0701113.

[29] K. Thepsuriya and A. Lewis, J. Cosmol. Astropart. Phys. **1**, 034 (2015), 1409.5066.

[30] E. Castorina, C. Carbone, J. Bel, E. Sefusatti, and K. Dolag, J. Cosmol. Astropart. Phys. **7**, 043 (2015), 1505.07148.

[31] S. Anselmi, D. López Nacir, and G. D. Starkman, Phys. Rev. D **92**, 084033 (2015), 1506.01000.

[32] S. Anselmi, S. Kumar, D. López Nacir, and G. D. Starkman, Phys. Rev. D **96**, 084001 (2017), 1706.01872.