Linear point standard ruler for galaxy survey data: Validation with mock catalogs

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, Ohio 44106-7079, USA

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

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

(Received 10 December 2017; published 25 July 2018)

Due to late-time nonlinearities, the location of the acoustic peak in the two-point galaxy correlation function is a redshift-dependent quantity, and 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 nonlinear 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 nonlinear 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 setup 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 of the Baryon Oscillation Spectroscopic Survey, for which the estimates of cosmic distances using the linear point have been presented by Anselmi *et al.* [Phys. Rev. Lett. **121**, 021302 (2018)].

DOI: 10.1103/PhysRevD.98.023527

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 [1–3]. 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 [4–6].

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 nonlinear dynamics of matter's gravitational clustering. Several studies, using both high-precision cosmological simulations and analytic models, have shown that nonlinearities distort the BAO pattern: smearing the BAO peak, lowering its amplitude and shifting

*stefano.anselmi@iap.fr

its position [7–10]. Therefore, peak-finding algorithms cannot be just blindly applied to the data to extract cosmic distance information, but rather the opposite—one should use cosmology-dependent fits of the full CF [11,12]. This would be a minor inconvenience if we knew how to predict the full nonlinear 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 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 nonlinearities and with the intent of "marginalizing over" the chosen fiducial cosmology [13,14]. This model template is then used to infer the cosmic distance from the statistical data analysis. Moreover, since nonlinearities suppress the amplitude of the BAO, the observed galaxy positions are adjusted, using approximate nonlinear 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 nonlinear effects [15]. 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), was suggested by some of us [16]. Its position, defined as the midpoint between the positions of the peak and dip in the monopole CF, is located at $\sim$95 Mpc/h in comoving units [16]. Using results from N-body simulations of ΛCDM models, it has been shown that the LP is insensitive to nonlinear 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 [16]. 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 modeling of the CF data.

Recently, we have presented [17] 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

$$
\begin{aligned}
D_V^{\mathrm{LP}}(\bar{z}_{\mathrm{LOWZ\text{-}DR12}} = 0.32) &= (1264 \pm 28)\ \mathrm{Mpc},\\
D_V^{\mathrm{LP}}(\bar{z}_{\mathrm{CMASS\text{-}DR12}} = 0.57) &= (2056 \pm 22)\ \mathrm{Mpc},
\end{aligned} \tag{1}
$$

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 Ref. [17] to the actual LOWZ and CMASS galaxy samples. To this end, we employ the Quick Particle Mesh (QPM) mock catalogs ("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 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 galaxy mock catalog, 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 Sec. 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, and we provide a checklist that the optimal fitting setup should pass to be validated. In Sec. III, we perform the previously introduced tests, discussing step by step the results of the analysis. In Sec. 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 catalogs, 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]

[1] https://www.sdss3.org/surveys/boss.php.

[2] Since we want to test the LP estimation procedure for different survey volumes, we do not focus on the final galaxy clustering analysis performed by the BOSS Collaboration [20]. We consider instead the CF analysis presented in Ref. [19] where the LOWZ and CMASS galaxy samples are taken into account.

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 catalog and its properties were built to match the mass function and large-scale bias of halos of high-resolution simulations. The halo catalog was then populated with galaxies using a halo occupation distribution (HOD) modeling, 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 catalog 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,

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

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

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

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.

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 Sec. III B. Second, the authors of Ref. [16] found that, in the BAO range of scales, the CF is nearly antisymmetric with respect to the LP. As we will show in Sec. 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 Sec. 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:

$$b_{\rm LP} = \bar{s}_{\rm LP} - s_{\rm LP}^{\rm true}. \qquad (4)$$

Here, $\bar{s}_{\rm LP}$ is the mean of the LP positions estimated from the mocks, and $s_{\rm LP}^{\rm true}$ 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 nonlinear clustering of matter [16]. As already mentioned in the Introduction, that nonlinear clustering systematically shifts the location of the LP relative to the linear CF up to 0.5% [16], 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 $\bar{s}_{\rm LP}$ and $s_{\rm LP}^{\rm true}$ equally.] We therefore take $s_{\rm LP}^{\rm true}$ to be the value estimated from the CF averaged over all the

[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 $y \equiv s/\text{constant}$ [12,16]. However, the procedure to extract the LP is the same whether the correlation function is expressed as a 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.

mocks. This has negligible cosmic-variance and sampling-variance error compared to the individual mocks but (deliberately) shares with them any potential systematic bias in $\bar{s}_{\rm LP}$.[5]

We would like to recall that measuring the LP position (or the BAO feature) is really a two-part process: detecting the LP and estimating its location. Indeed, given the finite volume of the mocks, the BAO feature in the CF might not be detected (by the chosen BAO estimator) in a given mock, due to cosmic variance. In our specific case, the polynomial estimator might fail to "detect" the peak and dip in the BAO range of scale; i.e., $d\xi_0^{\rm fit}/ds = 0$ could have no solutions. Clearly, only the mocks where the LP is detected can be used to estimate its error. To estimate it, we thus need to condition the analysis to the mocks in which the LP is detected (for each configuration of the polynomial-fit estimator, i.e., order of the polynomial, bin size, and range of scales). 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 data set, 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, the fraction of retained mocks [mock acceptance rate (MAR)] is $\gtrsim 80\%$ for LOWZ and $\gtrsim 90\%$ for CMASS. Notice that, since in the corresponding real data the LP is detected, the specific value of the MAR is not relevant for the present analysis [17]. As stated above, the bias and statistical error on the LP position are calculated using only the retained mocks. The rejected mocks contain no information on the LP position but contribute only to the false negative rate for LP detection. The need to separately minimize false negative rates, bias, and statistical error contributes to the design of all estimators of cosmological quantities, including other estimators of the BAO [26–28]. This is not widely discussed in the literature. Rather than looking at false negative rates, what is often done is imposing a threshold significance of BAO detection—typically taken to be 2σ.

Given the finite size of the mock samples, we correct the LP error budget according to Refs. [29,30]. Similarly, we follow Ref. [30] 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 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.

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

III. LINEAR POINT ESTIMATION TESTS

As mentioned in the Introduction, the advantage of the LP is that it is a geometric standard ruler on the BAO scale that is preserved by nonlinear 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 prereconstructed 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 $\chi^2_{\rm min}$ 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 $s_{\rm LP}$ 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 $(s_{\rm LP} - \bar{s}_{\rm LP})/\sigma_{s_{\rm LP}}$. The unit normal probability distribution function is overplotted. We perform 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 $s_{\rm LP}$ is Gaussian distributed. Therefore, we can

[5] In Ref. [16], 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 1.3% and 1.2% with respect to the linear-theory prediction respectively. This disagreement could be due to the approximate treatment of clustering in the modeling used to build 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., Ref. [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 nonlinear LP extraction, we can safely ignore this issue.

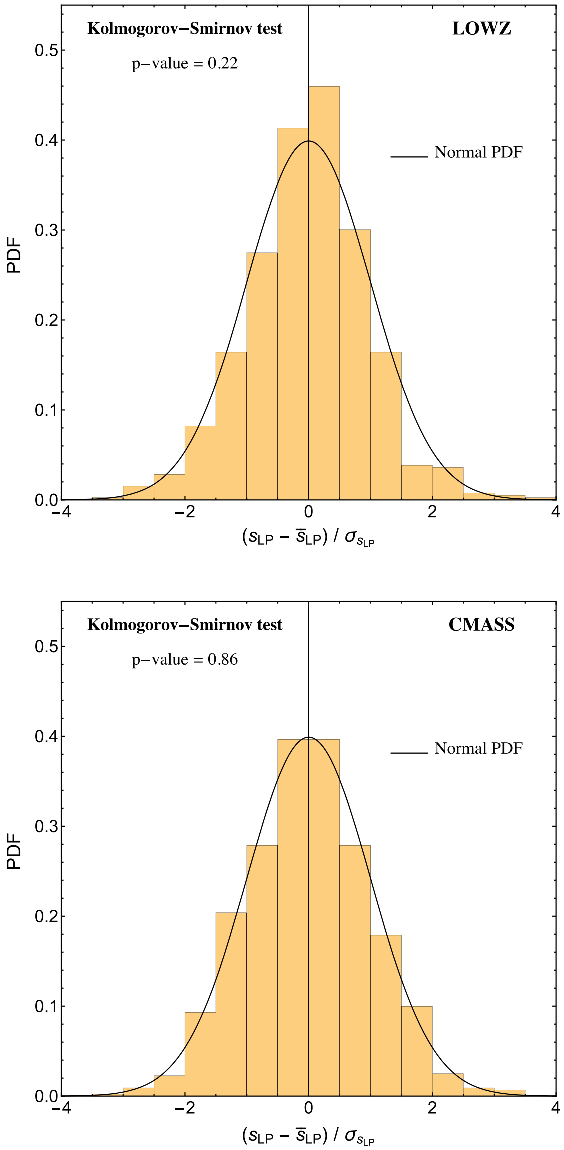

FIG. 1. Normalized histograms of the rescaled linear point positions recovered from the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogs. The unit normal probability distribution function is overplotted. The p values of the Kolmogorov-Smirnov test show that, for both catalogs, there is a reasonable probability that the LP values are drawn from a Gaussian distribution.

assign the usual Gaussian meaning to the rms of the LP distribution.[6]

The mean $s_{\rm LP}$ errors for the two simulated galaxy samples are $\sigma_{s_{\rm LP}}^{\rm LOWZ} = 2.4$ Mpc/h and $\sigma_{s_{\rm LP}}^{\rm CMASS} = 1.5$ Mpc/h. Therefore, the intrinsic 0.5% deviation of $s_{\rm LP}$ with respect to $s_{\rm LP}^{\rm lin}$ (found in Ref. [16]) is subdominant.

[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 [30], the two agree to better than 7%.

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.

Estimator test			
Polynomial	$b_{\rm LP}$	$\bar{\sigma}_{s_{\rm LP}}$	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

B. Optimal polynomial estimator

For LP estimators, we analyze polynomials of cubic or higher order.[7] For the LOWZ and CMASS galaxy surveys, we find that the cubic and the quartic polynomials return a LP systematic bias that is comparable to the statistical error budget. Therefore, since they are biased estimators, we do not analyze them any further in this manuscript.

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 $s_{\rm LP}$ 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 catalog, 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., $b_{\rm LP} < 0.2 \times \bar{\sigma}_{\rm LP}$. As expected, $\bar{\sigma}_{s_{\rm LP}^{\rm 5th}} < \bar{\sigma}_{s_{\rm LP}^{\rm 7th}}$, 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 diagonal line is an indication of the cosmic-variance error; indeed, due to cosmic variance, the LP value shows some scatter around its mean value. If two estimators would be completely correlated, that would be the only source of scatter. The presence of some scatter perpendicular to the solid line indicates a contribution from estimator error—the difference between the "true" value of the quantity being estimated (which we cannot know) and its estimated value. Quantitatively, Pearson's correlation coefficient r for the two samples, $r^{\rm LOWZ} = 0.67$ and $r^{\rm CMASS} = 0.63$, reveals that cosmic variance is the dominant error. We have checked that the combined use of the two estimators does not significantly reduce the statistical error compared with using only the

[7]Notice that lower-order polynomials are not LP estimators as they do not have both a maximum and a minimum.

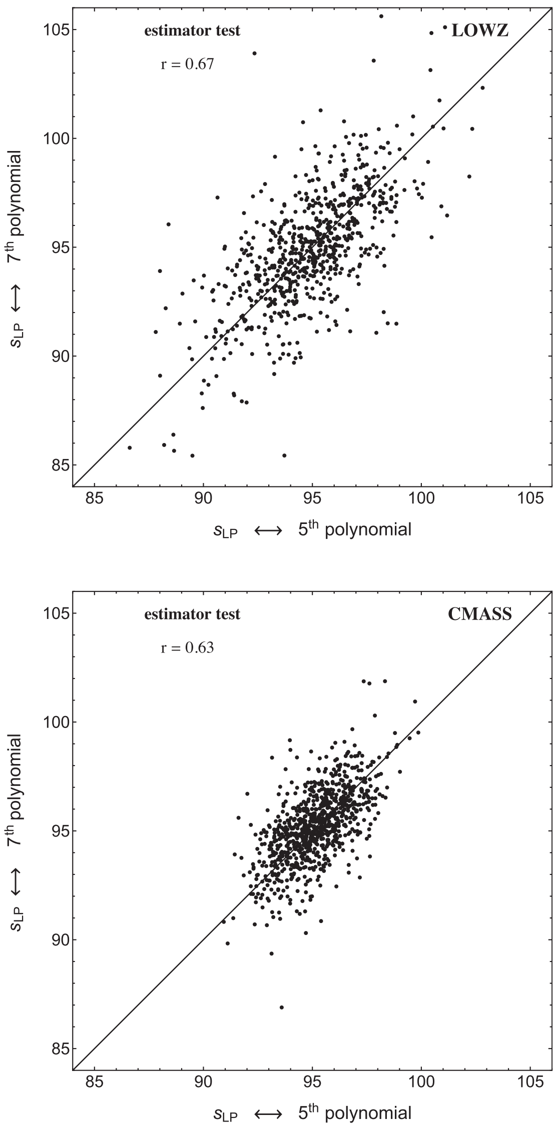

FIG. 2. Scatter in the LP *estimator* for the quintic *vs* the seventh-order polynomial fits to the CF of the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogs. The scatter along the continuous black line indicates the cosmic-variance error, while the scatter perpendicular to the line represents 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.

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 AIC_c [31]. We report its formula here for convenience (dropping an irrelevant additive constant),

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

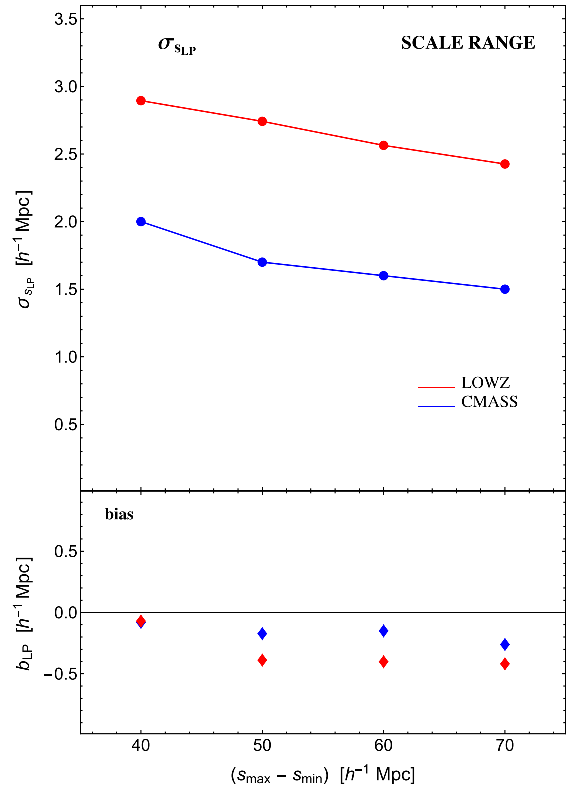

FIG. 3. LP-estimation error *(upper panel)* and bias *(lower panel)* as functions of the range of scales used to fit the CF. As we can see, the bias is always negligible, since $b_{LP} \leq 0.2 \times \bar{\sigma}_{s_{LP}}$.

where n was introduced in Eq. (2) and N is the number of points fit. The idea behind the 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 AIC_c value is selected. From Table I, we see that the smallest mean AIC_c belongs always to the quintic polynomial. This motivates its choice for the optimal setup.

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 antisymmetric with respect to the linear point over the BAO range of scales [16]. 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,

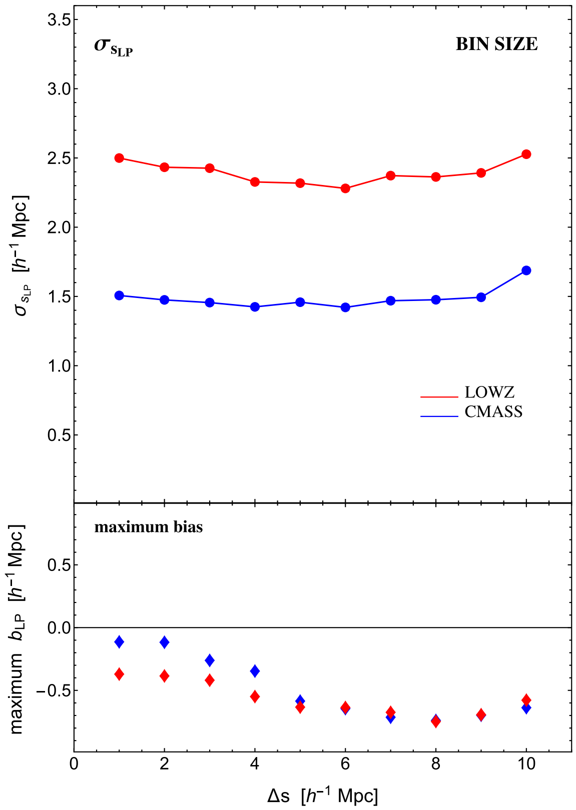

FIG. 4. LP-estimation error *(upper panel)* and bias *(lower panel)* as functions of the bin width. We see that the bias in the LP estimate induced by the choice of the binning of the CF is negligible (i.e., $b_{LP} \leq 0.2 \times \bar{\sigma}_{s_{LP}}$) for $\Delta s \leq 4$ Mpc/h.

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.

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 [17]. 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 is $(s_{max} - s_{min}) = 70$ Mpc/h.

D. Optimal bin size

The CF is measured in bins of finite width from data sets, 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 the CF in the BAO range of scales, and it introduces a dependence on the sampling choice.

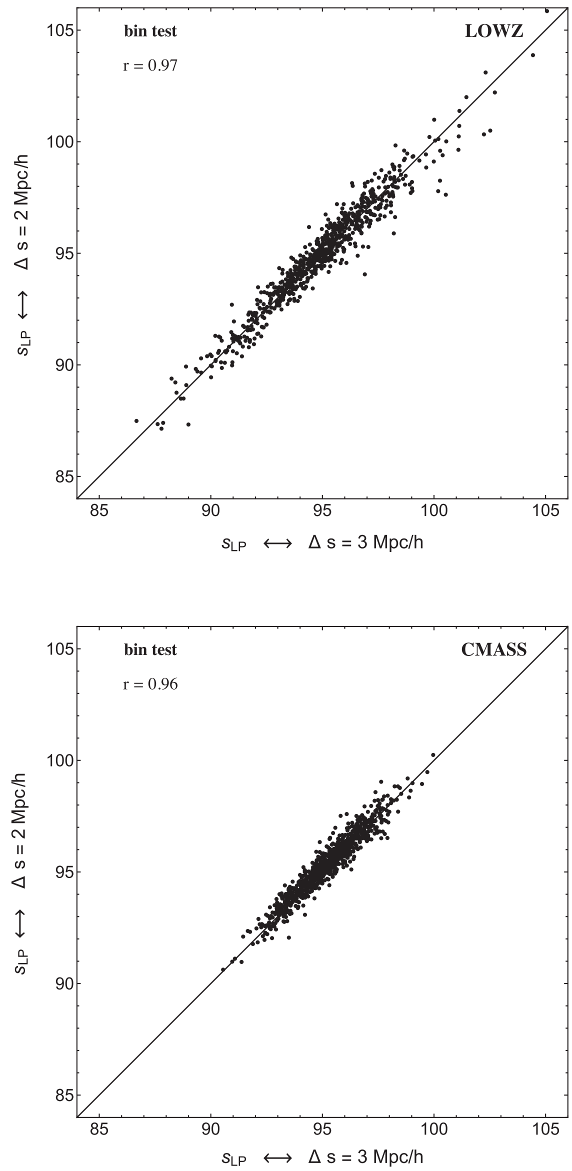

FIG. 5. LP-estimator scatter plot, for a quintic-polynomial fit to the CF, with bins of width $\Delta s = 2$ and 3 Mpc/h, for the LOWZ *(upper panel)* and CMASS *(lower panel)* mock catalogs.

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 us to reduce the covariance matrix noise [30], we choose $\Delta s = 3$ Mpc/h for the optimal setup.

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 nonlinearities 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 Ref. [17], 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 nonlinear physics that affects the galaxy correlation function at the BAO range of scales.

In this paper, we have determined the optimal setup to extract, by means of the LP, distance information from the BOSS-DR12 LOWZ and CMASS galaxy samples, justifying the methodology applied in Ref. [17]. This consists of 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 found that the peak and the dip are not detected in a fraction of the available mocks. In forecasting for a future survey, one would want to design both the survey and estimator to minimize the probability of such false negatives. Fortunately, with a judicious choice of estimator parameters, both LOWZ and CMASS data exhibit the needed peak and dip [17]. We consequently consistently condition our analysis to the mocks compatible with these observations, a practice followed by those characterizing other BAO estimators [26–28], although this is often not explicitly discussed.

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 nonlinearities [16]. Also worth investigating is the effect on the LP of massive neutrinos [32,33], 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 antisymmetric feature [16] is altered in candidate models (such as Quasidilaton Massive Gravity Theory [34,35]). Alternatively, one could construct a maximal-deviation test where, in the context of the concordance ΛCDM, the maximal allowed deviation from the predicted CF antisymmetry feature (in the BAO regime) is quantified.

ACKNOWLEDGMENTS

We are deeply grateful to Antonio Cuesta for providing us the correlation function QPM mock measurements, as well as for his careful reading of the manuscript and for his valuable suggestions. We thank Shadab Alam for discussions. The research leading to these results has received funding from the European Research Council under the European Community Seventh Framework Programme (FP7/2007-2013 Grant No. 279954) ERC-StG "EDECS." G. D. S. is supported by Department of Energy Grant No. DE-SC0009946 to the particle astrophysics theory group at CWRU. I. Z. is supported by National Science Foundation Grant No. AST-1612085.

[1] P. J. E. Peebles and J. T. Yu, Astrophys. J. **162**, 815 (1970).
[2] A. D. Sakharov, Sov. J. Exp. Theor. Phys. **22**, 241 (1966).
[3] J. Silk, Astrophys. J. **151**, 459 (1968).
[4] D. J. Eisenstein, W. Hu, and M. Tegmark, Astrophys. J. Lett. **504**, L57 (1998).
[5] B. A. Bassett and R. Hlozek, arXiv:0910.5224.
[6] D. J. Eisenstein *et al.*, Astrophys. J. **633**, 560 (2005).
[7] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **77**, 043525 (2008).
[8] D. J. Eisenstein, H.-J. Seo, and M. White, Astrophys. J. **664**, 660 (2007).
[9] T. Matsubara, Phys. Rev. D **77**, 063530 (2008).
[10] M. Crocce and R. Scoccimarro, Phys. Rev. D **77**, 023533 (2008).
[11] A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga, Mon. Not. R. Astron. Soc. **400**, 1643 (2009).
[12] A. G. Sanchez *et al.*, Mon. Not. R. Astron. Soc. **425**, 415 (2012).
[13] H.-J. Seo, E. R. Siegel, D. J. Eisenstein, and M. White, Astrophys. J. **686**, 13 (2008).
[14] X. Xu, N. Padmanabhan, D. J. Eisenstein, K. T. Mehta, and A. J. Cuesta, Mon. Not. R. Astron. Soc. **427**, 2146 (2012).
[15] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, Astrophys. J. **664**, 675 (2007).
[16] S. Anselmi, G. D. Starkman, and R. K. Sheth, Mon. Not. R. Astron. Soc. **455**, 2474 (2016).
[17] S. Anselmi, G. D. Starkman, P.-S. Corasaniti, R. K. Sheth, and I. Zehavi, Phys. Rev. Lett. **121**, 021302 (2018).
[18] M. White, J. L. Tinker, and C. K. McBride, Mon. Not. R. Astron. Soc. **437**, 2594 (2014).

[19] A. J. Cuesta *et al.*, Mon. Not. R. Astron. Soc. **457**, 1770 (2016).
[20] S. Alam *et al.*, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).
[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).
[24] M. Sinha, A. A. Berlind, C. K. McBride, R. Scoccimarro, J. A. Piscionere, and B. D. Wibking, arXiv: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).
[26] L. Anderson *et al.*, Mon. Not. R. Astron. Soc. **427**, 3435 (2012).
[27] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, Mon. Not. R. Astron. Soc. **449**, 835 (2015).
[28] J. E. Bautista *et al.*, arXiv:1712.08064.
[29] J. Hartlap, P. Simon, and P. Schneider, Astron. Astrophys. **464**, 399 (2007).
[30] W. J. Percival *et al.*, Mon. Not. R. Astron. Soc. **439**, 2531 (2014).
[31] A. R. Liddle, Mon. Not. R. Astron. Soc. **377**, L74 (2007).
[32] K. Thepsuriya and A. Lewis, J. Cosmol. Astropart. Phys. 01 (2015) 034.
[33] E. Castorina, C. Carbone, J. Bel, E. Sefusatti, and K. Dolag, J. Cosmol. Astropart. Phys. 07 (2015) 043.
[34] S. Anselmi, D. L. Nacir, and G. D. Starkman, Phys. Rev. D **92**, 084033 (2015).
[35] S. Anselmi, S. Kumar, D. L. Nacir, and G. D. Starkman, Phys. Rev. D **96**, 084001 (2017).