# Cosmological inference from an emulator based halo model. I. Validation tests with HSC and SDSS mock catalogs

Hironao Miyatake[1,2,3,*] Yosuke Kobayashi,[3,4] Masahiro Takada[3,†] Takahiro Nishimichi[5,3] Masato Shirasaki,[6,7] Sunao Sugiyama[3,4] Ryuichi Takahashi[8] Ken Osato[9] Surhud More[10,3] and Youngsoo Park[3]

[1]*Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan*
[2]*Division of Particle and Astrophysical Science, Graduate School of Science,*
*Nagoya University, Nagoya 464-8602, Japan*
[3]*Kavli Institute for the Physics and Mathematics of the Universe (WPI),*
*The University of Tokyo Institutes for Advanced Study (UTIAS),*
*The University of Tokyo, Chiba 277-8583, Japan*
[4]*Physics Department, The University of Tokyo, Bunkyo, Tokyo 113-0031, Japan*
[5]*Center for Gravitational Physics, Yukawa Institute for Theoretical Physics,*
*Kyoto University, Kyoto 606-8502, Japan*
[6]*National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan*
[7]*The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan*
[8]*Faculty of Science and Technology, Hirosaki University, 3 Bunkyo-cho, Hirosaki,*
*Aomori 036-8561, Japan*
[9]*Institut d'Astrophysique de Paris, Sorbonne Université, CNRS,*
*UMR 7095, 98bis boulevard Arago, Paris 75014, France*
[10]*The Inter-University Centre for Astronomy and Astrophysics,*
*Post bag 4, Ganeshkhind, Pune 411007, India*

We present validation tests of emulator-based halo model method for cosmological parameter inference, assuming hypothetical measurements of the projected correlation function of galaxies, $w_p(R)$, and the galaxy-galaxy weak lensing, $\Delta\Sigma(R)$, from the spectroscopic SDSS galaxies and the Hyper Suprime-Cam Year 1 (HSC-Y1) galaxies. To do this, we use DARK EMULATOR developed in Nishimichi *et al.* based on an ensemble of $N$-body simulations, which is an emulation package enabling a fast, accurate computation of halo clustering quantities (halo mass function, halo autocorrelation and halo-matter cross-correlation) for flat-geometry cold dark matter cosmologies. Adopting the halo occupation distribution, the emulator allows us to obtain model predictions of $\Delta\Sigma$ and $w_p$ for the SDSS-like galaxies at a few CPU seconds for an input set of parameters. We present performance and validation of the method by carrying out Markov chain Monte Carlo analyses of the mock signals measured from a variety of mock catalogs that mimic the SDSS and HSC-Y1 galaxies. We show that the halo model method can recover the underlying true cosmological parameters to within the 68% credible interval, except for the mocks including the assembly bias effect (although we consider the unrealistically large amplitude of assembly bias effect). Even for the assembly bias mock, we demonstrate that the cosmological parameters can be recovered if the analysis is restricted to scales $R \gtrsim 10 \ h^{-1}$ Mpc (i.e., if the information on the average mass of halos hosting SDSS galaxies inherent in the 1-halo term of $\Delta\Sigma$ is not included). We also show that, by using a single population of source galaxies to infer the relative strengths of $\Delta\Sigma$ for multiple lens samples at different redshifts, the joint probes method allows for self-calibration of photometric redshift errors and multiplicative shear bias. Thus we conclude that the emulator-based halo model method can be safely applied to the HSC-Y1 dataset, achieving a precision of $\sigma(S_8) \simeq 0.04$ after marginalization over nuisance parameters such as the galaxy-halo connection parameters and the photo-$z$ error parameter, and our method is complementary to methods based on perturbation theory.

## I. INTRODUCTION

Wide-area imaging galaxy surveys offer exciting opportunities to address the fundamental questions in cosmology such as the nature of dark matter and the origin of cosmic

*hironao.miyatake@iar.nagoya-u.ac.jp
†masahiro.takada@ipmu.jp

083519-1

acceleration [1]. The current-generation imaging surveys such as the Subaru Hyper Suprime-Cam [2] (HSC) [3–5], the Dark Energy Survey [6] (DES) [7], and the Kilo-Degree Survey (KiDS) [8] have succeeded in using accurate measurements of weak gravitational lensing effects to obtain tight constraints on cosmological parameters. Interestingly, the cosmological model inferred from these large-scale structure probes show the so-called $\sigma_8$ or $S_8$ tension (e.g., [4,9]) compared with cosmological models inferred from the *Planck* cosmic microwave background (CMB) measurement [10], perhaps indicating a signature beyond the standard cosmological model, i.e., the flat-geometry Lambda-dominated cold dark matter ($\Lambda$CDM) model (e.g., [9]). Upcoming galaxy surveys such as the Subaru Prime Focus Spectrograph [11,12], the Dark Energy Spectrograph Instrument [13], the VRO Legacy Survey of Space and Time [14], the ESA Euclid [15], and the NASA Roman Space Telescope [16,17] are expected to deliver a decisive conclusion on the possible tension and also revolutionize our understanding of the Universe.

Main challenges of large-scale structure probes lie in uncertainty in galaxy bias, which refers to unknown relation between the distributions of matter and galaxies in large-scale structure [18,19]. Since physical processes inherent in formation and evolution of galaxies are still difficult to accurately model from the first principles, we need both observational and theoretical approaches to tackle the galaxy bias uncertainty in order for us to obtain "unbiased" and "precise" estimation of the underlying cosmological parameters from large-scale structure observables.

For observational approach, joint probes cosmology offers a promising way to mitigate the impact of galaxy bias uncertainty on cosmology inference [7,8,20–26]. In particular, galaxy-galaxy weak lensing, obtained by cross-correlating positions of foreground (lens) galaxies with shapes of background galaxies, can be used to infer the average mass distribution around lens galaxies. A combination of galaxy-galaxy weak lensing with autocorrelation function of the same sample of galaxies as the lens galaxies can be used to observationally disentangle the galaxy bias and the correlation function of the underlying matter distribution from the measured clustering signal of galaxies.

On theory side, there are mainly two empirical approaches to tackle the galaxy bias uncertainty. First one is a model based on perturbation theory (PT) of large-scale structure [27,28]. As long as only the large-scale information of clustering observables in the linear or quasinonlinear regime is used and nuisance parameters to model galaxy bias are introduced, such a PT-based method is expected to serve as an "accurate" theoretical template of galaxy clustering [26,29–31]. Accuracy here refers to the fact that a PT model can reproduce the observed clustering correlation of galaxies by varying the bias parameters, down to a certain scale still in the quasinonlinear regime, where PT is valid. An advantage of this method is that the model can be used for any type of galaxies, because PT is formulated based on properties of gravity and primordial fluctuations [32] and the free bias parameters absorb large-scale clustering properties of galaxies irrespective of galaxy types. A price to pay is that the method breaks down at scales below a certain nonlinear scale, and cannot be used to extract cosmological information from the small-scale clustering signals, which generally carry higher signal-to-noise ratios than in the large scales. A further refined method that separates short-scale physics such as bias from large-scale cosmological information of interest, so-called effective field theory of large scale structure [33], has been emerging as a method that can be applicable to small scales, although an application to galaxy-galaxy lensing and clustering measurements has yet to be made.

An alternative theoretical method is the halo model approach [34–37]. Halos are places where galaxies likely form, and clustering properties of halos are relatively well understood, on both analytical approach and *N*-body simulations [38]. Then an empirical model such as the halo occupation distribution (HOD) method [39,40] can be used to connect halos to galaxies. An advantage of this method is that it would allow one to use the small-scale information in cosmology inference, thereby yielding tighter constraints on cosmological parameters. However, a danger is that, if the model is not sufficiently accurate nor flexible enough to capture the complicated galaxy-scale physics, the method might lead to a significant bias in cosmological parameters, more than the statistical credible interval. A worst-case scenario is that one might claim a wrong cosmology, e.g., a time-varying dark energy model, from a given dataset due to the inaccurate theoretical templates.

Hence, the purpose of this paper is to assess performance and limitation of the halo model method for cosmology inference. To do this, we use the DARK EMULATOR developed in Nishimichi *et al.* [41], which enables a fast, accurate computation of halo clustering quantities (halo mass function, halo autocorrelation function and halo-matter cross-correlation) for an input set of cosmological parameters within flat-geometry *w*CDM framework with adiabatic Gaussian initial conditions. The DARK EMULATOR is particularly useful for our halo model approach, as it enables accurate predictions for the clustering quantities well into the nonlinear regime and thereby allows us to make robust use of the cosmological information from small scales. We combine DARK EMULATOR with the HOD model to make model predictions of the projected correlation functions of galaxies, $w_p(R)$ and the galaxy-galaxy weak lensing, $\Delta\Sigma(R)$, that mimic those measured from the spectroscopic SDSS DR11 galaxies [42] and the HSC Year 1 (HSC-Y1) galaxies [43]. More precisely, we consider mock galaxies of LOWZ and CMASS galaxies in the redshift range $0.15 \lesssim z \lesssim 0.7$ for the spectroscopic galaxies in the $w_p$ measurement and for the lens sample in $\Delta\Sigma$, and then consider mock galaxies of the deep HSC-Y1 data for the background galaxy sample in $\Delta\Sigma$. We use a variety of mock catalogs for the SDSS galaxies to assess performance

and limitation of the halo model method for cosmology inference, including a mock where we implement an extreme version of galaxy assembly bias [44]. Here we quantify the performance by studying whether the halo model method can recover the cosmological parameters employed in the mock catalogs, after marginalization over the galaxy-halo connection parameters—hereafter we will often refer to this exercise as cosmology challenges. Moreover, following Oguri and Takada [21], we assess the ability of our method to perform self-calibration of photometric redshift errors and multiplicative bias error for the joint probes cosmology. Those errors are among the most important, systematic errors for weak lensing cosmology, and we show below that the method allows for an accurate self-calibration of the errors, i.e., a robust estimation of cosmological parameters mitigating the impact of the systematic errors. This paper gives a validation of the halo model methodology for cosmology inference that we are planning to carry out with the actual SDSS and HSC datasets [45]. The results of this paper are also compared to those of the companion paper, Sugiyama *et al.* [26], which used a perturbation theory inspired method and was validated against the same mock catalogs.

The structure of this paper is as follows. In Sec. II we briefly review the halo model for $w_p$ and $\Delta\Sigma$ and discuss the methodology of how cosmological parameters can be estimated from the clustering observables. In Sec. III we describe details of $N$-body simulations, the halo catalogs, the mock catalogs of SDSS and HSC galaxies, and DARK EMULATOR. In Sec. IV we describe our strategy for assessing performance and limitation of the halo model method, i.e., cosmology challenges done from the comparison of the halo model with the mock signals. In Sec. V, we show the main results of this paper. Section VI is devoted to conclusion and discussion. Throughout this paper, unless otherwise stated, we employ the flat-geometry *Planck* cosmology [46] as a target cosmology for cosmology challenges and as a fiducial cosmology when we compute cosmological dependences of observables. The model is characterized by $\Omega_m = 0.3156$ (the present-day matter density parameter), $h = 0.672$ and $\sigma_8 = 0.831$, respectively, and we adopt the units of $c = 1$ for the speed of light.

## II. THEORY

In this paper we focus on two observables that are obtained from imaging and spectroscopic data of galaxies in the overlapping regions of the sky. One is the galaxy-galaxy weak lensing that can be measured by stacking shapes of background galaxies around a sample of foreground lensing galaxies. Here we assume that the background galaxy sample is taken from the Subaru HSC images, while the foreground lensing galaxies are from the SDSS spectroscopic galaxies. The other is the projected correlation function of the spectroscopic galaxies that are from the same population of galaxies used as lens (foreground) galaxies in the galaxy-galaxy weak lensing.

### A. Observables: Galaxy-galaxy weak lensing and projected autocorrelation function

Galaxy-galaxy weak lensing [24,47] probes the averaged excess surface mass density profile, $\Delta\Sigma$, around the lensing (foreground) galaxies that is given in terms of the surface mass density profile, $\Sigma_{gm}(R)$, as

$$\Delta\Sigma(R; z_l) = \langle\Sigma_{gm}\rangle(< R) - \Sigma_{gm}(R),$$
$$= \Sigma_{cr}(z_l, z_s)\gamma_+(R)|_{R=\chi_l\Delta\theta}, \quad (1)$$

where $\gamma_+$ is the average tangential shear of background galaxies in the circular annulus of projected centric radius $R$ from the foreground galaxies, and $\chi_l$ is the comoving angular diameter distance to each foreground galaxy. Note that background galaxy shapes are averaged over all the pairs of foreground-background galaxies in the same projected separation $R$, not the angular separation $\Delta\theta$, even when the lensing galaxies have a redshift distribution. $\langle\Sigma_{gm}\rangle(< R)$ is the average surface mass density within a circular aperture of radius $R$, defined as $\langle\Sigma_{gm}\rangle(< R) \equiv 1/(\pi R^2) \int_0^R 2\pi R' dR' \Sigma_{gm}(R')$. $\Sigma_{cr}$ is the critical surface mass density that describes a lensing efficiency for pairs of foreground/background galaxies as a function of their redshifts, and is defined as

$$\Sigma_{cr}(z_l, z_s) \equiv \frac{\chi_s(z_s)}{4\pi G \chi_{ls}(z_l, z_s)\chi_l(z_l)(1 + z_l)}, \quad (2)$$

where $\chi_s$ and $\chi_{ls}$ are the comoving angular diameter distances to each source galaxy and between source and lens galaxies in each pair, respectively. The factor $(1 + z_l)$ is due to our use of the comoving coordinates. In practice we need to take into account the redshift distributions of lens and source galaxies that are straightforward to include, e.g., following the method in Ref. [24].

The surface mass density profile around lensing galaxies, $\Sigma_{gm}(R)$, is given in terms of the three-dimensional galaxy-matter cross-correlation function as

$$\Sigma_{gm}(R; z_l) = \bar{\rho}_{m0} \int_{-\infty}^{\infty} d\pi \xi_{gm}(\sqrt{\pi^2 + R^2}; z_l), \quad (3)$$

where $\bar{\rho}_{m0}$ is the mean mass density today, $\xi_{gm}(r)$ is the three-dimensional cross-correlation function between the galaxies and matter, and $\pi$ is the separation along the line-of-sight direction, and $R$ is the projected separation perpendicular to the line-of-sight direction. Here we ignored the contribution from the background mean mass density because it is not relevant for weak lensing observables [compared to Eq. (3) in Ref. [26] ]. The Fourier-transformed counterpart of $\xi_{gm}(r)$ is the galaxy-matter cross power spectrum, $P_{gm}(k)$, defined as

$$\xi_{gm}(r; z_l) \equiv \int_0^{\infty} \frac{k^2 dk}{2\pi^2} P_{gm}(k; z_l) j_0(kr), \quad (4)$$

where $j_0(x)$ is the zeroth-order spherical Bessel function. Hereafter we omit the redshift of the lensing galaxies, $z_l$, in the argument of the correlation functions for notational simplicity. Throughout this paper we assume a distant observer approximation for computations of projected correlation functions. In other words, we ignore the effects of curved sky for definitions of separations along or perpendicular to the line-of-sight direction. The excess surface mass density profile is given in terms of the galaxy-matter cross power spectrum as

$$\Delta\Sigma_{\mathrm{gm}}(R) \equiv \langle\Sigma_{\mathrm{gm}}\rangle(< R) - \Sigma_{\mathrm{gm}}(R),$$
$$= \bar{\rho}_{\mathrm{m0}} \int_0^\infty \frac{k^2 \mathrm{d}k}{2\pi^2} P_{\mathrm{gm}}(k) J_2(kR), \qquad (5)$$

where $J_2(x)$ is the second-order Bessel function.

To obtain the theoretical template of $\Delta\Sigma(R)$ for a given cosmological model within the flat $\Lambda$CDM framework, we use DARK EMULATOR, developed in Ref. [41], which enables accurate and fast computations of halo statistical quantities; the halo mass function, the halo-matter cross-correlation function and the halo-autocorrelation function as a function of redshift, halo masses, and separations for an input cosmological model. As demonstrated in Ref. [41], the emulator outputs the predictions to better than a few percent in the fractional amplitude over scales of separations we are interested in.

However, the emulator does not take into account the effects of baryonic physics, and this is a limitation we should keep in mind. Nevertheless we do not think this limitation causes a catastrophic failure of our approach due to the following reasons. The baryonic physics is local in the sense that it affects the matter distribution at scales smaller than a maximum scale, denoted as $R_*$. For example, even if we consider a violent effect of the AGN feedbacks, it would affect the mass distribution around halos up to a few Mpc at maximum. As nicely discussed in Refs. [48,49], we can safely consider that the baryonic effect causes a "redistribution" of matter at $r \lesssim R_*$ around galaxies, and does not alter the mass distribution at $r \gtrsim R_*$ ($r$ is the three-dimensional radius from the center of the galaxy). In other words, even in the presence of the baryonic physics, the mass conservation at $r \lesssim R_*$ holds. Keeping this in mind, we can say that DARK EMULATOR, as designed, can accurately model the matter distribution at $r \gtrsim R_*$ around the host halos of galaxies, and also satisfies the mass conservation to within $r \simeq R_*$, even though DARK EMULATOR ceases to accurately predict the mass profile at $r \lesssim R_*$.

Based on the above consideration, we can rewrite the lensing profile [Eq. (5)] as

$$\Delta\Sigma(R) = \langle\Sigma_{\mathrm{gm}}\rangle(< R) - \Sigma_{\mathrm{gm}}(R),$$
$$= \frac{M(< R_*)}{\pi R^2} + \frac{2}{R^2}\int_{R_*}^R R'\mathrm{d}R'\Sigma_{\mathrm{gm}}(R') - \Sigma_{\mathrm{gm}}(R). \quad (6)$$

In the second line on the rhs, we used the relation $M(< R_*) \equiv \int_0^{R_*} 2\pi R'\mathrm{d}R'\Sigma_{\mathrm{gm}}(R')$, where $M(< R_*)$ is the mass interior of a circular aperture of radius $R_*$. The second term is the contribution over the range of $[R_*, R]$ to the average mass density. As long as we focus on the lensing profiles at $R \geq R_*$, DARK EMULATOR can accurately predict the lensing profile, including the interior mass at the aperture of $R \simeq R_*$. On the other hand, we need to introduce various nuisance parameters to model the lensing "profile" at $R \leq R_*$, if the information is included, and then marginalize over the parameters when estimating the interior mass $M(< R_*)$ and performing cosmology inference.

Another clustering observable we use is the projected autocorrelation function for spectroscopic galaxy sample that is the same sample used as lens (foreground) galaxies in the galaxy-galaxy weak lensing measurement. The projected correlation function is defined by a line-of-sight projection of the three-dimensional autocorrelation function of galaxies, $\xi_{\mathrm{gg}}(r)$, as

$$w_{\mathrm{p}}(R) \equiv 2\int_0^{\pi_{\max}} \mathrm{d}\Pi\,\xi_{\mathrm{gg}}(\sqrt{R^2 + \Pi^2}), \qquad (7)$$

where $\pi_{\max}$ is the length of the line-of-sight projection. Note that the projected correlation function has the unit of $[h^{-1}\,\mathrm{Mpc}]$. Throughout this paper, unless explicitly stated, we employ $\pi_{\max} = 100\ h^{-1}\,\mathrm{Mpc}$ as our default choice. Also notice that $\xi_{\mathrm{gg}}(r)$ is given in terms of the autopower spectrum of the galaxy number density field, $P_{\mathrm{gg}}$, as

$$\xi_{\mathrm{gg}}(r) = \int_0^\infty \frac{k^2 \mathrm{d}k}{2\pi^2} P_{\mathrm{gg}}(k) j_0(kr). \qquad (8)$$

The projected correlation function is not sensitive to the redshift-space distortion (RSD) effect due to peculiar velocities of galaxies, if a sufficiently large projection length ($\pi_{\max}$) is taken (see Fig. 6 in Ref. [50]). The RSD effect itself is a useful cosmological probe, but its use requires an accurate modeling [44,51,52], which is not straightforward. Hence, the projected correlation function makes it somewhat easier to compare with theory in a cosmological analysis. In the following, we ignore the RSD effect in most cases of our cosmology challenges, but will separately discuss the impact of the RSD effect in parameter estimation.

### B. Theoretical template: Dark Emulator implementation of halo model

In this section we describe details of the halo model implementation to make model predictions for the observables $\Delta\Sigma(R)$ and $w_{\mathrm{p}}(R)$. As can be found around Eqs. (5) and (7), we need the real-space cross-power spectrum of galaxies and matter $P_{\mathrm{gm}}(k; z)$, and the real-space autopower spectrum of galaxies, $P_{\mathrm{gg}}(k; z)$, as a function of cosmological models to compute the observables.

### 1. Halo occupation distribution

In the halo model we assume that all matter is associated with halos, and the correlation function of matter is given by the contributions from pairs of matter in the same halo and those in two different halos, which are referred to as the 1- and 2-halo terms, respectively. To connect halos to galaxies in a given sample, we employ the HOD [35,37,39]. The HOD model gives the mean number of central and satellite galaxies in halos of mass $M$ as

$$\langle N \rangle(M) = \langle N_c \rangle(M) + \langle N_s \rangle(M), \qquad (9)$$

where $\langle \rangle(M)$ denotes the average of a quantity for halos of mass $M$.

We employ the mean HOD for central galaxies, given as

$$\langle N_c \rangle(M) = \frac{1}{2} \left[ 1 + \mathrm{erf} \left( \frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right], \qquad (10)$$

where $\mathrm{erf}(x)$ is the error function and $M_{\min}$ and $\sigma_{\log M}$ are model parameters.

For the mean HOD of satellite galaxies, we employ the following form:

$$\langle N_s \rangle(M) \equiv \langle N_c \rangle(M) \lambda_s(M) = \langle N_c \rangle(M) \left( \frac{M - \kappa M_{\min}}{M_1} \right)^\alpha, \qquad (11)$$

where $\kappa$, $M_1$, and $\alpha$ are model parameters, and we have introduced the notation $\lambda_s(M) = [(M - \kappa M_{\min})/M_1]^\alpha$. For our default prescription we assume that satellite galaxies reside only in a halo that already hosts a central galaxy. This means $N_c = 1$ for such halos. Then we assume that the number of satellite galaxies in halos of mass $M$ follows the Poisson distribution with mean $\lambda_s(M)$; $\mathrm{Prob}(N_s) = (\lambda_s)^{N_s} \exp(-\lambda_s)/N_s!$ for halos that host a central galaxy inside, and $\mathrm{Prob}(N_s) = \delta^K_{N_s,0}$ for halos that do not host a central galaxy, where $\delta^K_{ij}$ is the Kronecker delta function. Under these assumptions, the mean number of galaxy pairs living in the same halo with mass $M$, which is relevant for the 1-halo term, can be computed as

$$
\begin{aligned}
\langle N(N-1) \rangle &= \langle N_c \rangle \langle N(N-1) \rangle|_{N_c=1} \\
&\quad + (1 - \langle N_c \rangle) \langle N(N-1) \rangle|_{N_c=0}, \\
&= \langle N_c \rangle [\langle (N_c + N_s)(N_c + N_s - 1) \rangle]_{N_c=1}, \\
&= \langle N_c \rangle [\langle N_s^2 \rangle + \langle N_s \rangle], \\
&= \langle N_c \rangle [2\lambda_s + \lambda_s^2], \qquad (12)
\end{aligned}
$$

where we have used the fact $N = 0$ for halos which have no central galaxy for our default prescription.

We have five model parameters, $\{M_{\min}, \sigma_{\log M}, \kappa, M_1, \alpha\}$, to characterize the central and satellite HODs in total for each galaxy sample for a given cosmological model.

Once the HOD model is given, the mean number density of galaxies in a sample is given as

$$\bar{n}_g = \int dM \frac{dn_h}{dM} [\langle N_c \rangle(M) + \langle N_s \rangle(M)], \qquad (13)$$

where $dn_h/dM$ is the halo mass function which gives the mean number density of halos in the mass range $[M, M + dM]$. In the following we use "blue" fonts to denote a quantity that can be supplied by DARK EMULATOR. In the above case we use DARK EMULATOR to compute the halo mass function for $w$CDM cosmology, while we input the HOD model, as described, to compute the galaxy number density.

### 2. Galaxy-galaxy weak lensing profile

As we described above, the galaxy-galaxy weak lensing arises from the cross-correlation of (spectroscopic) lensing galaxies with the surrounding matter distribution in large-scale structure, $\xi_{gm}(r; z_l)$. Hence we need to model $\xi_{gm}(r)$ as a function of the parameters of galaxy-halo connection and cosmological parameters.

Since the Fourier space gives a somewhat simpler form of the expression, we mainly write down the power spectrum that is the Fourier-transformed counterpart of the two-point correlation function, here $\xi_{gm}$. Under the halo model approach, the cross-power spectrum of galaxies and matter is given as

$$
\begin{aligned}
P_{gm}(k) = \frac{1}{\bar{n}_g} \int dM \frac{dn_h}{dM} [\langle N_c \rangle(M) + \langle N_s \rangle(M) \tilde{u}_s(k; M, z)] \\
\times P_{hm}(k), \qquad (14)
\end{aligned}
$$

where $P_{hm}(k)$ is the halo-matter cross-power spectrum that can be computed by DARK EMULATOR as a function of cosmological model within $w$CDM cosmologies. The quantity $\tilde{u}_s(k; M)$ is the Fourier transform of the averaged radial profile of satellite galaxies in host halos of $M$ at redshift $z$, which we need to specify. Note that all the quantities in the above equation are evaluated at the lens redshift $z_l$, but we omit to denote in the argument of each function for notational simplicity. DARK EMULATOR was built to calibrate the halo-matter cross-correlation by measuring the averaged mass profile around halos with mass $M$ in $N$-body simulation outputs [41]. By construction, the halo-matter cross-correlation satisfies the mass conservation around halos. We also note that DARK EMULATOR already includes both the 1- and 2-halo term contributions in $P_{hm}$, which correspond to the cross-correlations of halos with matter in the same halo and the surrounding matter outside the halo, respectively.

More exactly speaking, DARK EMULATOR outputs $\xi_{\rm hm}(r; M)$ for an input set of parameters (halo mass, separation, and cosmological parameters), which is the Fourier transform of $P_{\rm hm}(k; M)$. In order to obtain $\xi_{\rm gm}$ or $\Delta\Sigma$ for the assumed model we extensively use the publicly available FFTLOG [53] code to perform the Hankel transforms when going back and forth between real and Fourier space [see Eqs. (4) and (5)].

For the radial profile of satellite galaxies, we assume that satellite galaxies follow a Navarro-Frenk-White (NFW) profile [54] that is an approximated model of the averaged dark matter profile in halos. To compute the NFW profile as a function of halo mass, redshift, and cosmological model, we need to specify the scaling relation of mass concentration with halo mass in the halo matter profile. For this we employ the fitting formula given in Ref. [55]; more exactly, we use the publicly available PYTHON code, "COLOSSUS" [56,57], to compute the concentration parameter for a given set of parameters (halo mass, redshift, and cosmological parameters).

As for the default model of the galaxy-halo connection, we consider neither the off-centering effect nor the incompleteness effect of galaxies, where the latter describes a possibility that some fraction of even very massive halos might not host a SDSS-like galaxy due to an incomplete selection after the specific color and magnitude cuts (for details of the model, see [25]). Rather than introducing additional nuisance parameters to model these effects, we employ a minimum halo model as our baseline model. However, we use different types of mock catalogs of SDSS-like galaxies including the off-centering effects and the incompleteness effect, and then will use the mock catalogs to validate and assess the performance of the baseline method. If our baseline model can recover the underlying cosmological parameters, we claim that the method is validated. If the method shows any failure to recover the cosmological parameters, we will start to introduce more parameters. However, if the cosmological constraints turn to be sensitive to such details of a treatment of galaxies inside the halo, then this is a sign of the failure or limitation of the halo model based method, because such a small-scale distribution of galaxies is very difficult to accurately model due to complexities of physical processes inherent in formation and evolution of galaxies. Hence, rather than including such model parameters, we employ the minimum halo model to assess its performance.

The upper panel of Fig. 1 shows how different terms of $P_{\rm gm}$ contributes to the $\Delta\Sigma$ profile for the SDSS-like galaxies at $z = 0.251$ as we will describe in more detail. Even though satellite galaxies tend to reside in massive halos, all the curves have a similar shape ($R$ dependence) in large $R$ bins, $R \gtrsim 10\ h^{-1}$ Mpc, for a fixed cosmology. All the small-scale physics involved in the galaxy-halo connection affects $\Delta\Sigma$ at $R \lesssim 10\ h^{-1}$ Mpc. Note that, due to the nonlocal nature of $\Delta\Sigma$, the small-scale physics affects $\Delta\Sigma$
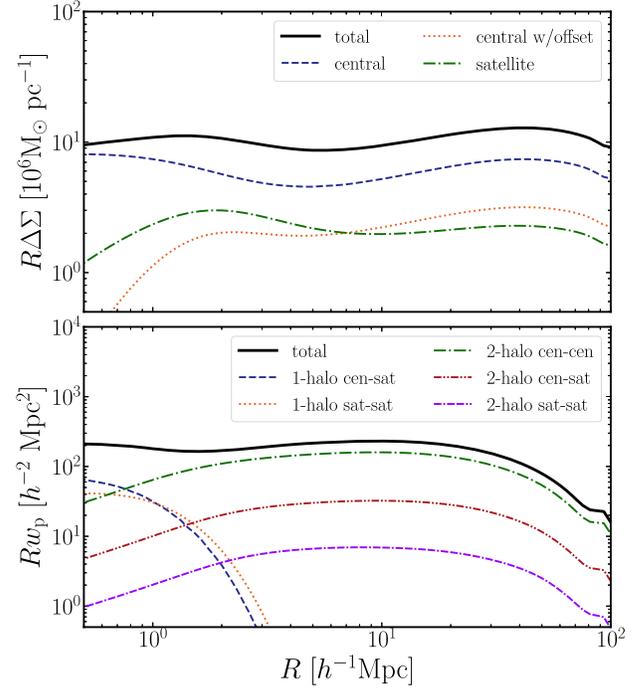


FIG. 1. Break down of different contributions to the galaxy weak lensing profile [$\Delta\Sigma(R)$; upper panel] and the projected correlation function of galaxies ($w_{\rm p}$; lower) for the SDSS LOWZ-like galaxies at $z = 0.251$, which are computed using DARK EMULATOR and our model ingredients of the galaxy-halo connection for the *Planck* cosmology. For illustrative purpose we multiply each observable by $R$ so that their dynamic range ($y$ axis) becomes narrower. Upper panel: the dashed line shows a contribution arising from the cross-correlation of central galaxies with the surrounding matter distribution, while the dot-dashed line denotes a contribution from the cross-correlation of satellite galaxies with matter. For comparison, we also show how a possible off-centering of central galaxies affects the lensing profile, although we do not include this effect in the theoretical templates. Here, as a working example, we consider the off-centering parameters $p_{\rm off} = 0.3$ and $\mathcal{R}_{\rm off} = 0.4$ irrespective of halo mass, where $p_{\rm off}$ models a fraction of off-centered central galaxies in halos of mass $M$, while $\mathcal{R}$ is the off-centering radius relative to the scale radius of NFW profile of halos. The upper solid lines is the total power. Lower: the similar break down of different contributions to the projected correlation function of the galaxies, for the same model in the upper panel.

up to relatively large scales, compared to a virial radius of massive halos. Also importantly, the integrated lensing signal up to a few Mpc scales gives an estimate of the average halo mass, which has a close tie to the large-scale amplitudes of $\Delta\Sigma$ and $w_{\rm p}$ via the scaling relation of halo bias with halo mass (see below). For comparison, the figure also shows the impact of off-centering effects of central galaxies on $\Delta\Sigma$, assuming that some fraction of central galaxies might be offset from the true halo center as a result of the assembly history of galaxies in their host halos [58,59].

### 3. Projected correlation function of galaxies

To model $w_{\rm p}$ we need to model the real-space correlation function of galaxies, $\xi_{\rm gg}$, or the Fourier transform $P_{\rm gg}$ for an input set of parameters. The autopower spectrum of galaxies in a sample is decomposed into the two contributions, the 1- and 2-halo terms, and those are given within the halo model framework as

$$P_{\rm gg}(k) \equiv P_{\rm gg}^{\rm 1h}(k) + P_{\rm gg}^{\rm 2h}(k) \tag{15}$$

with

$$P_{\rm gg}^{\rm 1h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \int {\rm d}M \frac{{\rm d}n_{\rm h}}{{\rm d}M} \langle N_{\rm c} \rangle (M) [2\lambda_{\rm s}(M)\tilde{u}_{\rm s}(k;M) + \lambda_{\rm s}(M)^2 \tilde{u}_{\rm s}(k;M)^2],$$

$$P_{\rm gg}^{\rm 2h}(k) = \frac{1}{\bar{n}_{\rm g}^2} \left[ \int {\rm d}M \frac{{\rm d}n_{\rm h}}{{\rm d}M} \langle N_{\rm c} \rangle (M) \{1 + \lambda_{\rm s}(M)\tilde{u}_{\rm s}(k;M)\} \right]$$
$$\times \left[ \int {\rm d}M' \frac{{\rm d}n_{\rm h}}{{\rm d}M'} \langle N_{\rm c} \rangle (M') \{1 + \lambda_{\rm s}(M')\tilde{u}_{\rm s}(k;M')\} \right] P_{\rm hh}(k;M,M'), \tag{16}$$

where $P_{\rm hh}(k;M,M')$ is the power spectrum between two halo samples with masses $M$ and $M'$. DARK EMULATOR outputs the real-space correlation function of halos, $\xi_{\rm hh}(r;M,M')$, that is the Fourier transform of $P_{\rm hh}$. Similarly to the case of $\Delta\Sigma$, we use the FFTLOG code to perform the Fourier transform to obtain the prediction of $\xi_{\rm gg}(r)$ [see Eq. (8)] and then perform the line-of-sight integral to obtain the prediction of $w_{\rm p}(R)$ [see Eq. (7)]. We note that, strictly speaking, our standard halo model implementation of the 1-halo term behaves as a shot noise like term of $k^0(= {\rm const.})$ at the limit of $k \to 0$, and this gives a subtle violation of the mass and momentum conservation [32]. One could modify the halo model at very small $k$ to enforce the conservation laws, as done in Ref. [60], but we found that our treatment, practically, gives a sufficiently accurate model prediction over scales of separations that we are interested in, for the SDSS-like galaxies [51] (see also [61] for a similar discussion). As given by Eq. (7), we employ $\pi_{\rm max} = 100~h^{-1}~{\rm Mpc}$, as our default choice to integrate $\xi_{\rm gg}(r = \sqrt{R^2 + \pi^2})$ to obtain the model prediction for $w_{\rm p}(R)$.

The lower panel of Fig. 1 shows different contributions in the projected correlation function $w_{\rm p}$ for the *Planck* cosmology, assuming the same model as in the upper panel. As can be found from Eq. (16), the 1-halo term arises from the central-satellite and satellite-satellite correlations, while the 2-halo term is from the correlations of central-central, central-satellite and satellite-satellite galaxies in different halos, respectively. The figure clearly shows that all the contributions to the 1-halo term are confined to small scales, at $R \lesssim$ a few Mpc, a virial radius of massive halos, reflecting the local nature of $w_{\rm p}$ in contrast to $\Delta\Sigma$. All the different contributions to the 2-halo term have a similar shape ($R$ dependence); the different terms differ from each other only by a multiplicative factor for a fixed cosmology. This means that the shape of the 2-halo term is not sensitive to details of the galaxy-halo connection. The features around $R \simeq 90~h^{-1}~{\rm Mpc}$ are the BAO features, which appear at smaller scales than the BAO scale of $100~h^{-1}~{\rm Mpc}$ in the three-dimensional correlation function, due to the line-of-sight projection. However, we note that we will not include the BAO information in the parameter estimation in the following; we will use the $w_{\rm p}$ information up to $R \simeq 30~h^{-1}~{\rm Mpc}$ as our default choice.

## C. Joint probes cosmology: A mitigation method of galaxy bias uncertainty

For convenience of our discussion let us define the cross-correlation coefficient function for halo correlation function:

$$r_{\rm hm}(r) \equiv \frac{\xi_{\rm hm}(r)}{[\xi_{\rm hh}(r)\xi_{\rm mm}(r)]^{1/2}}. \tag{17}$$

As shown in Fig. 31 of Ref. [41], DARK EMULATOR predicts $r_{\rm hm} \simeq 1$ on $r \gtrsim 10~h^{-1}~{\rm Mpc}$, the scale greater than a typical size of massive halos, and $r_{\rm hm}$ is close to unity within 5% or so even at the intermediate scales a few Mpc $< r \lesssim 10~h^{-1}~{\rm Mpc}$. As stressed in our companion paper [26], the real-space observables have an advantage that all variants due to the galaxy-halo connection are confined to small scales, $r \lesssim 10~h^{-1}~{\rm Mpc}$, and the cross-correlation function on large scales satisfy $r_{\rm hm} \simeq 1$ (see also [62] for a similar discussion). This is not necessarily true in Fourier space, because the variations on small scales become extended in Fourier space due to the nature of Fourier transform. The simplest example is the shot noise; the shot noise affects the real-space correlations at zero separation, while it behaves like a white noise and affects the power spectrum over all scales ($k$ bins).

It is also constructive to discuss an asymptotic behavior of the halo correlation functions, $\xi_{\rm gm}$ and $\xi_{\rm gg}$, on large scales. At scales much greater than the nonlinear scale, $r \gg R_*$, the galaxy-matter cross correlation has an asymptotic behavior given as z

$$\xi_{\rm gm}(r)\overline{[r \gg R_*]} \frac{1}{\bar{n}_{\rm g}} \int {\rm d}M \frac{{\rm d}n_{\rm h}}{{\rm d}M} \langle N_{\rm g} \rangle (M)\xi_{\rm hm}(r;M)$$

$$\simeq b_{\rm eff}\xi_{\rm mm}(r), \tag{18}$$

where we have used $\tilde{u}_s(k; M) \to 1$ for $k \ll 1/R_*$, and we defined the effective bias parameter, at a sufficiently large separation satisfying $r \gg R_*$, as

$$b_{\rm eff} \equiv \frac{1}{\bar{n}_g \xi_{\rm mm}(r)} \int dM \frac{dn_h}{dM} \langle N_g \rangle \xi_{\rm hm}(r; M). \quad (19)$$

In addition, as can be found from Eq. (3), $\Delta\Sigma$, which has a dimension of $[hM_\odot \, {\rm Mpc}^{-2}]$, has an additional dependence on $\bar{\rho}_{m0}$

$$\Delta\Sigma \propto \bar{\rho}_{m0} \xi_{\rm gm} \propto \Omega_m b_{\rm eff} \xi_{\rm mm}. \quad (20)$$

Note that the dependence of $h$ is not relevant because all the quantities are measured in units where the dependence of $h$ is factorized out.

Similarly the autocorrelation function of galaxies is found to have an asymptotic behavior at the limit of large scales, $r \gg R_*$, as

$$\xi_{\rm gg}(r) \overrightarrow{[r \gg R_*]} \frac{1}{\bar{n}_g^2} \left[ \int dM \frac{dn_h}{dM} \langle N_g \rangle (M) \right]$$
$$\times \left[ \int dM' \frac{dn_h}{dM'} \langle N_g \rangle (M') \right] \xi_{\rm hh}(r; M, M') \simeq b_{\rm eff}^2 \xi_{\rm mm}(r).$$
$$(21)$$

This is not an exact relation, but we found that the relation $r_{\rm gg} \equiv \xi_{\rm gm}/[\xi_{\rm gg}\xi_{\rm mm}]^{1/2} \simeq 1$ holds for scales of $r \gg R_*$, as shown in Fig. 32 of Ref. [41]. Thus the combination of $\xi_{\rm gm}$ and $\xi_{\rm gg}$, which are inferred from $\Delta\Sigma$ and $w_p$, respectively, can be used to infer the underlying matter correlation function $\xi_{\rm mm}$, and in turn we can use it to extract cosmological information. Furthermore, as implied from Eq. (6), an amplitude of $\Delta\Sigma$ at scales around a transition scale between the 1- and 2-halo terms gives an estimate of the average mass of halos hosting galaxies:

$$\Delta\Sigma_{\rm gm}(R)|_{R\simeq{\rm a \ few \ Mpc}} \to \bar{M}_h. \quad (22)$$

In turn this can put a strong constraint on the halo bias, $b_{\rm eff}$, via the scaling relation of halo bias with halo mass (or the dependence of $\xi_{\rm hm}$ and $\xi_{\rm hh}$ amplitudes on halo mass). Hence, combining the small- and large-scale information of $\Delta\Sigma$ with $w_p$ helps break degeneracies between cosmological parameters and the galaxy bias, and then gives useful constraints on cosmological parameters. This is the basic picture of how the joint probes cosmology using $\Delta\Sigma$ and $w_p$ can constrain cosmological parameters. However, if the scaling relation of the large-scale galaxy bias amplitude with the average halo mass is broken, then this method does not work. This would be the case for the assembly bias effect, where the galaxy bias depends on secondary parameter(s) related to the assembly history of host halos, in addition to halo mass.
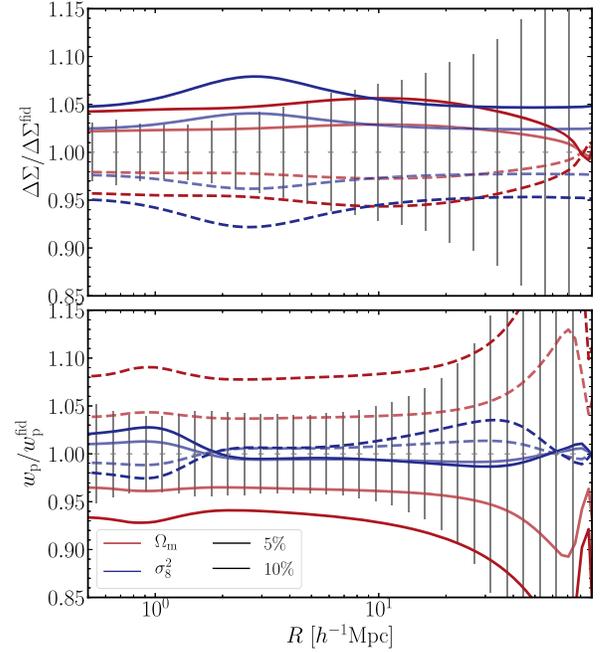


FIG. 2. Dependences of $\Delta\Sigma$ and $w_p$ on cosmological parameters, $\sigma_8$ (blue lines) and $\Omega_m$ (red), computed using the DARK EMULATOR based halo model. Here we consider the HOD parameters for the LOWZ sample at $z = 0.251$. Here we consider fractional changes of $\sigma_8^2$ (not $\sigma_8$) or $\Omega_m$ by $\pm 5\%$ or $\pm 10\%$, respectively, where the other parameters are fixed to their fiducial values. The solid and dashed, respectively, lines show the fractional changes in $\Delta\Sigma$ or $w_p$ relative to that for the fiducial model when $\sigma_8^2$ or $\Omega_m$ is changed to a positive or negative side from its fiducial value. The error bars around unity denote $1\sigma$ statistical errors that are computed from the diagonal terms of the covariance matrix expected for the SDSS and HSC-Y1 data, although the neighboring bins are highly correlated with each other.

In Fig. 2, we study how $\Delta\Sigma$ and $w_p$ depend on the cosmological parameters, $\Omega_m$ and $\sigma_8$, because large-scale structure probes are most sensitive to these parameters. Here we consider the expected signals of $\Delta\Sigma$ and $w_p$ for the SDSS LOWZ-like galaxies at $z = 0.251$ (as we will describe in detail in Sec. III B) which are computed using DARK EMULATOR based on the halo model. We note that, due to the nonlocal nature of $\Delta\Sigma$, the lensing profile at a particular scale, say $R_0$, is sensitive to the matter-galaxy cross-correlation at $r \sim R_0/2$. Note that we employ $R_{\rm cut} = 2$ and $3 \, h^{-1} \, {\rm Mpc}$ for the scale cuts of $\Delta\Sigma$ and $w_p$ as our default choice, respectively, and will use the information of $\Delta\Sigma$ and $w_p$ at scales greater than the scale cuts for the cosmology challenges.

Let us first consider the dependence of $\sigma_8$. There are two competing effects in the change of our target observables originating from that of $\sigma_8$ for models with a fixed $\Omega_m$. First, an increase of $\sigma_8$ boosts the amplitude of $\xi_{\rm mm}$ by definition. Second, it leads to a decrease of halo bias

($b_{\mathrm{eff}}$) for massive halos hosting SDSS-like galaxies ($\sim 10^{13}\ h^{-1}\,M_\odot$), because a model with higher $\sigma_8$ leads to more evolved large-scale structure at an observed redshift, thus the abundance of massive halos gets increased at the redshift, and such halos become less biased tracers, resulting in a lowered bias amplitude compared to that for the fiducial model. For $\Delta\Sigma$, the first effect is more significant; an increase of $\sigma_8$ leads to an increase of the amplitude of $\Delta\Sigma$ over all the scales we consider. For $w_{\mathrm{p}}$, these competing effects almost cancel out on large scales, $R \gtrsim$ a few Mpc in the 2-halo term, and the increase of $\sigma_8$ does not largely change the amplitude on the large scales. Nevertheless we should emphasize that the change of $\sigma_8$ causes a scale-dependent modification in these observables, especially $w_{\mathrm{p}}$, in contrast to the linear theory that predicts a constant (scale-independent) shift in the ratio at these large scales (also see Fig. 6 in [26] for the similar discussion). We checked that the scale-dependent changes are from a combination of the effects of $\sigma_8$ changes on matter clustering ($\xi_{\mathrm{mm}}$) and the halo bias function (see Fig. 19 in Appendix A). These effects are automatically built in the DARK EMULATOR predictions. Also note that the positive- and negative-side changes of $\sigma_8$ cause an asymmetric change in $w_{\mathrm{p}}$ at large scales. On the other hand, the

1-halo term amplitude of $w_{\mathrm{p}}$ is boosted because the increase of $\sigma_8$ leads to the increased abundance of massive halos, more satellite galaxies reside in such halos, and their correlations add. Thus the change of $\sigma_8$ leads to characteristic modifications of the amplitude and scale dependence in $\Delta\Sigma$ and $w_{\mathrm{p}}$. The notable features around $R \sim 90\ h^{-1}$ Mpc are due to the effect on BAO features [63], but this is not relevant for our results because we use the information of $\Delta\Sigma$ and $w_{\mathrm{p}}$ on scales up to $30\ h^{-1}$ Mpc for the fiducial choice, so do not include the BAO information in parameter estimation.

Next we discuss the results for the change of $\Omega_{\mathrm{m}}$, for a fixed $\sigma_8$. First, an increase of $\Omega_{\mathrm{m}}$ leads to a faster evolution of clustering growth for models with a fixed $\sigma_8$ (the fixed normalization today), so this leads to the smaller amplitude of $\xi_{\mathrm{mm}}$ at this redshift ($z = 0.251$). However, recalling that $\Delta\Sigma \propto \Omega_{\mathrm{m}} \langle \gamma_+ \rangle \sim \Omega_{\mathrm{m}} \delta_{revfourrmm}$, an increase of $\Omega_{\mathrm{m}}$ leads to higher amplitudes in $\Delta\Sigma$ due to the prefactor, while it leads to smaller amplitudes in $w_{\mathrm{p}}$. In addition, the change of $\Omega_{\mathrm{m}}$ causes a scale-dependent modification in both $\Delta\Sigma$ and $w_{\mathrm{p}}$.

In Fig. 3 we study how a change in each of the HOD parameters alters $\Delta\Sigma$ and $w_{\mathrm{p}}$ for the *Planck* cosmology. Here we employ the same HOD parameters for the LOWZ
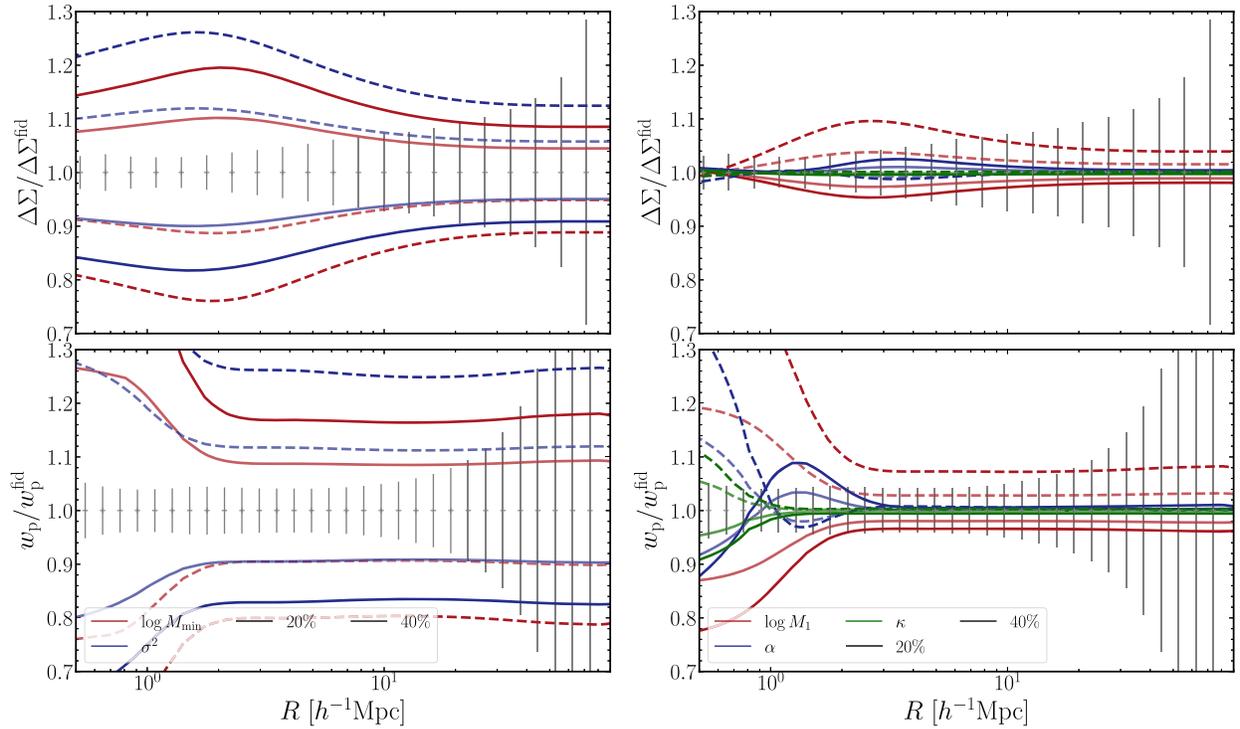


FIG. 3. Dependence of $\Delta\Sigma$ and $w_{\mathrm{p}}$ on changes to HOD parameters can be seen in the top and the bottom panels, respectively. Here the fiducial HOD parameters are chosen to resemble the SDSS LOWZ-like galaxy sample at $z = 0.251$ for the *Planck* cosmology. When one HOD parameter is changed, the other parameters are fixed to their fiducial values. Solid line shows the result for an increase (positive-side change) of each HOD parameter from its fiducial value, while dashed line is the result for the decrease. From light- to dark-color lines show the results for the fractional change of each HOD parameter by $\pm 20\%$ and $\pm 40\%$, respectively. The left panel shows the dependences of the central HOD parameters, $M_{\mathrm{min}}$ and $\sigma_{\log M}$, while the right panel shows the results for the satellite HOD parameters, $M_1, \kappa$ and $\alpha$. The error bars are the same in the previous figure.

galaxies as in Fig. 2, and vary only one HOD parameter for each result, where other parameters are fixed to their fiducial values. The figure shows that, when each of the central HOD parameters is varied, it causes a significant change over all the scales including both the 1- and 2-halo terms, because the change in the parameter leads to a change in the mean halo mass. More precisely, an increase in $M_{\mathrm{min}}$ or a decrease in $\sigma_{\log M}$ leads to an increase in the mean halo mass, leading to the increased amplitudes in $\Delta\Sigma$ and $w_{\mathrm{p}}$. The fractional changes in $\Delta\Sigma$ and $w_{\mathrm{p}}$ at large separations in the 2-halo term regime are roughly given as $\delta w_{\mathrm{p}}/w_{\mathrm{p}} \simeq 2\delta\Delta\Sigma/\Delta\Sigma$, reflecting the facts $\xi_{\mathrm{gg}} \simeq b_{\mathrm{eff}}^2 \xi_{\mathrm{mm}}$ and $\xi_{\mathrm{gm}} \simeq b_{\mathrm{eff}} \xi_{\mathrm{mm}}$ at the large scale limit as we discussed above. In addition, it is clear that the 1-halo term amplitude of $\Delta\Sigma$ is sensitive to the central HOD parameters, physically to the mean mass of halos hosting the galaxies in a sample. Hence, combining the 2-halo term amplitudes of $\Delta\Sigma$ and $w_{\mathrm{p}}$ allows one to break degeneracies between the bias parameter and other parameters, and then adding the 1-halo term information of $\Delta\Sigma$ can constrain the HOD parameters and then tighten the determination of the large-scale bias. The satellite HOD parameters change the relative contribution of massive halos to $\Delta\Sigma$ and $w_{\mathrm{p}}$ as satellite galaxies preferentially reside in massive halos. In particular, the parameter $M_1$, which determines the amplitude of the satellite HOD, causes a decent change in the large-scale amplitudes of $\Delta\Sigma$ and $w_{\mathrm{p}}$. The effects of other parameters are relatively mild.

Comparing Figs. 2 and 3 manifests that the cosmological parameters and the HOD parameters lead to different changes in $\Delta\Sigma$ and $w_{\mathrm{p}}$. Hence combining these observables in the 1- and 2-halo term regimes allows for an efficient determination of the cosmological parameters, by breaking the parameter degeneracies. These figures nicely illustrate the complementarity of $\Delta\Sigma$ and $w_{\mathrm{p}}$, which is the main focus of this paper.

### D. Observational effects: Geometrical cosmology dependence, photo-z, multiplicative shear bias, and RSD

In this section we discuss the four observational effects, i.e., the geometrical dependence of the observables on cosmology, photometric redshift errors of source galaxies, multiplicative shear bias, and redshift-space distortion effect. When comparing the measured signals with the model templates, we need to include these effects in the model templates, and here we describe how to do this.

#### 1. Geometrical cosmology dependence

The observables we consider in this paper are $\Delta\Sigma(R)$ and $w_{\mathrm{p}}(R)$, which are different from other possible choices of the lensing and clustering observables such as $\gamma_+(\theta)$ and $w_{\mathrm{p}}(\theta)$. Measurements of $\Delta\Sigma$ and $w_{\mathrm{p}}$ require an observer to assume a reference cosmology to perform a correction of the lensing efficiency $\Sigma_{\mathrm{cr}}(z_{\mathrm{l}}, z_{\mathrm{s}})$ [see Eq. (1)] as well as the conversion of angular scales ($\theta$) and redshift differences to the projected separation ($R$) and radial separation ($\pi$). However, the assumed cosmology generally differs from the true underlying cosmology, and this dependence needs to be taken into account. We use the method in Ref. [64] to include this effect.

For a flat-geometry $\Lambda$CDM model, the relevant parameter is only $\Omega_{\mathrm{m}}$ (or $\Omega_{\mathrm{de}}$ that is the density parameter of dark energy), because it affects the angular and radial distances (the lensing efficiency also depends on the combination of angular diameter distances and the overall factor of $\Omega_{\mathrm{m}}$). Note that the dependence on $h$ is taken out from the observables because all the quantities are measured in units of $h^{-1}$ Mpc or $hM_\odot$ Mpc$^{-2}$ for $w_{\mathrm{p}}$ and $\Delta\Sigma$, respectively. The $\Omega_{\mathrm{m}}$ dependences of $\Delta\Sigma$ and $w_{\mathrm{p}}$ through this effect turn out to be very small, but add a slight sensitivity of $\Omega_{\mathrm{m}}$ in the theoretical templates.

#### 2. Photometric redshift errors

Photometric redshift errors of source galaxies are one of the most serious systematic errors in the weak lensing measurements. An accuracy of photometric redshifts (hereafter often simply photo-z), delivered from a set of broadband filters (*grizy* in the HSC data), is limited, and can never be perfect, compared to spectroscopic redshifts, although the photo-z accuracies are calibrated using the COSMOS catalog [65]. For lens redshift we assume spectroscopic redshifts as we will focus on the spectroscopic SDSS-like galaxies. Here we discuss how we can treat a possible uncertainty of source redshifts in the observable $\Delta\Sigma$. In this paper we use the method proposed in Oguri and Takada [21]. In this method we use a "single" population of source galaxies, selected based on the photo-z, and then use the relative strengths of $\Delta\Sigma$ for multiple lens samples at different redshifts to calibrate the photo-z uncertainty of source galaxies in a statistical sense.

In the presence of a redshift distribution of source galaxies, an estimator of $\Delta\Sigma(R)$ from the measured ellipticity component of each source galaxy is given, e.g., by Eq. (11) in Ref. [24], as

$$\widehat{\Delta\Sigma}(R) = \frac{1}{2\mathcal{R}} \frac{\sum_{\mathrm{l,s}} w_{\mathrm{ls}} [\langle \Sigma_{\mathrm{cr}}^{-1} \rangle_{\mathrm{ls}}]^{-1} e_{\mathrm{ls}+}}{\sum_{\mathrm{l,s}} w_{\mathrm{ls}}}, \qquad (23)$$

where $\mathcal{R}$ is the "responsivity" that is needed to convert the measured galaxy ellipticity, defined in terms of $(a^2 - b^2)/(a^2 + b^2)$ ($a$, $b$ is the major, minor axes when the galaxy shape is approximated by an ellipse), to the lensing shear, defined in terms of $(a - b)/(a + b)$ [23,43,66]; $e_{\mathrm{s}+}$ is the tangential ellipticity component of the sth source galaxy with respect to the lth lensing galaxy; $w_{\mathrm{ls}}$ is the weight (see below). The summation $\sum_{\mathrm{l,s}}$ runs over all pairs of source and lens galaxies that are in a given bin of projected

separation, $R = \chi(z_1)\Delta\theta$. The measured lensing signal is for the effective redshift of lens galaxies, $\bar{z}_l$.

In a case that we have only the posterior distribution of photometric redshift for each source galaxy, the "effective" lensing efficiency needs to be estimated as

$$\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{\mathrm{ls}} \equiv \int_0^\infty \mathrm{d}z_{\mathrm{s}}\Sigma_{\mathrm{cr}}^{-1}(z_l, z_{\mathrm{s}})p(z_{\mathrm{s}}), \qquad (24)$$

where $p(z_{\mathrm{s}})$ is the posterior distribution of redshift for the sth source galaxy that satisfies the normalization condition $\int_0^\infty \mathrm{d}z_{\mathrm{s}}p(z_{\mathrm{s}}) = 1$. Note that we set $\Sigma_{\mathrm{cr}}^{-1}(z_l, z_{\mathrm{s}}) = 0$ when $z_{\mathrm{s}} < z_l$. We employ the weight $w_{\mathrm{ls}}$, motivated by the inverse-variance weight, to have a higher signal-to-noise ratio for the weak lensing measurement [67–69]:

$$w_{\mathrm{ls}} = \frac{[\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{(l,s)}]^2}{e_{\mathrm{rms}}^2 + \sigma_e^2}, \qquad (25)$$

where $e_{\mathrm{rms}}$ is the rms intrinsic ellipticity per component and $\sigma_e$ is the measurement error of ellipticity for the sth source galaxy.

Thus, in an estimator of $\Delta\Sigma$, the quantities $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{\mathrm{ls}}$ and $w_{\mathrm{ls}}$ depend on photometric redshifts of source galaxies via $p(z_{\mathrm{s}})$ for individual source galaxies. If the estimated redshift distribution of source galaxies is systematically offset from the underlying true distribution, then the measured $\Delta\Sigma$ has a systematic bias (offset) from the true one (even if the assumed cosmological model for the $\Delta\Sigma$ estimation is the true cosmology). Following the method in Refs. [21,70], we take into account the systematic offset by introducing additional nuisance parameter, $\Delta z_{\mathrm{ph}}$, to shift the posterior distribution of source redshift for all source galaxies as

$$p(z_{\mathrm{s}}) \to p(z_{\mathrm{s}} + \Delta z_{\mathrm{ph}}). \qquad (26)$$

The above method is a statistical method to correct for a possible residual difference between the true redshift distribution of source galaxy and the assumed distribution. Hence the parameter $\Delta z_{\mathrm{ph}}$ is designed to model a change in $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle$ due to the difference, and does not necessarily model a change in the mean and/or shape of source redshift distribution. This can be found from the following simple example. Consider the case that the lens galaxies are at a single redshift of $z_l = 0.5$ and the photo-$z$ posterior is approximated by a Gaussian with mean $z_{\mathrm{s}} = 1.3$ and $\sigma(z_{\mathrm{s}}) = 0.1$. Then we compute the effective source redshift $z_{s,\mathrm{eff}}$ that provides the same $\Sigma_{\mathrm{cr}}^{-1}$ as the average $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{\mathrm{ls}}$. However, if the true photo-$z$ posterior distribution is given by the Gaussian distribution with the same mean, but wider width $\sigma(z_{\mathrm{s}}) = 0.2$, we find the effective redshift is changed by $\Delta z_{s,\mathrm{eff}} = -0.025$, which is smaller than the prior width $\sigma(\Delta z_{\mathrm{ph}}) = 0.04$ or $0.1$, used in our cosmology analysis. Thus even if the shift in Eq. (26) cannot model a change in

the width or shape of the photo-$z$ posterior distribution, the effect can be safely captured by a change in $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle$, therefore the parameter $\Delta z_{\mathrm{ph}}$, as long as the change is covered by the prior width of $\Delta z_{\mathrm{ph}}$.

With this shift in the photo-$z$ posterior, we can repeat the computations of $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{\mathrm{ls}}$ and $w_{\mathrm{ls}}$ for the pairs of source and lens galaxies in actual SDSS and HSC-Y1 datasets. We found that the lensing profile after shifting the source redshift distribution is well approximated by the following multiplicative form as

$$\widehat{\Delta\Sigma}(R; \bar{z}_l, \Delta z_{\mathrm{ph}}) \simeq f_{\mathrm{ph}}(R; \bar{z}_l, \Delta z_{\mathrm{ph}})\widehat{\Delta\Sigma}(R; \bar{z}_l, \Delta z_{\mathrm{ph}} = 0), \qquad (27)$$

where $f_{\mathrm{ph}}(R; \bar{z}_l, \Delta z_{\mathrm{ph}})$ is the multiplicative factor to model the effect of systematic photo-$z$ error, and $\bar{z}_l$ is the mean of lens redshifts. Here we stress, by notation "^," that the above correction is made for the measured $\Delta\Sigma$. By using a single population of source galaxies, we notice that the shift $\Delta z_{\mathrm{ph}}$ leads to changes in the amplitudes of $\Delta\Sigma$ for each of the multiple lens samples (LOWZ and the two subsamples of CMASS, divided into two redshift bins) depending on the lens redshift ($\bar{z}_l$). Conversely, we can use the relative variations in the $\Delta\Sigma$ amplitudes at different lens redshifts to calibrate out $\Delta z_{\mathrm{ph}}$, simultaneously with cosmological parameter estimation. This is a self-calibration method of photo-$z$ errors as proposed in Oguri and Takada [21]. In the above equation, we explicitly include the $R$ dependence in the calibration factor, which could arise because the shift could change relative contributions of different lens-source pairs to the lensing efficiency $\langle\Sigma_{\mathrm{cr}}^{-1}\rangle_{\mathrm{ls}}$. However, the $R$ dependence, albeit very weak, appears only for the highest-redshift lens sample (CMASS2 in our sample, as we will define later).

For our method, we multiply the inverse of the calibration factor in Eq. (27) with the theoretical template of $\Delta\Sigma$, rather than correcting for the measurement, for an assumed $\Delta z_{\mathrm{ph}}$:

$$\Delta\Sigma(R) \to \Delta\Sigma(R)/f_{\mathrm{ph}}(R; \Delta z_{\mathrm{ph}}). \qquad (28)$$

We then include $\Delta z_{\mathrm{ph}}$ as additional nuisance parameter when carrying out the parameter inference. This is better because we will use the same covariance matrix in our parameter inference, which allows for apple-to-apple comparison of the performance for different setups (because this method does not change the error bars of $\Delta\Sigma$ in each $R$ bin).

### 3. Multiplicative shear errors

An accurate weak lensing measurement requires an exquisite, accurate characterization of individual galaxy shapes. This is not straightforward [43], and an imperfect shape measurement leaves a residual systematic error in the

weak lensing measurements. Systematic errors in shape measurements are usually modeled by "multiplicative" and "additive" biases in the measured galaxy ellipticities, given as $\gamma \rightarrow (1 + m)\gamma + c$, where $m$ and $c$ are the multiplicative and additive bias parameters, respectively [71]. Since the spatial positions of lensing galaxies on the sky are considered random with respect to the positions of source galaxies, the galaxy-galaxy weak lensing is not affected by the additive bias [43]. Hence, in the presence of the multiplicative shear bias, we modify the theoretical template as

$$\Delta\Sigma(R) \rightarrow (1 + m)\Delta\Sigma(R; m = 0). \quad (29)$$

By working on the single sample of source galaxies, we can employ the single $m$ parameter for all the lensing profiles at multiple lens redshifts. This is another advantage of the method of Ref. [21]. This is a good approximation as long as source galaxies are well separated from lens redshifts; this is the case as long as we can ignore a contamination of source galaxies into lens redshifts due to photo-$z$ errors. To make this method work, in actual data, we will employ a conservative cut of source galaxy redshifts to ensure that source galaxies are well separated from lens redshifts [45].

### 4. Redshift-space distortion

To model the projected correlation function of galaxies, we ignored the RSD effect that is caused by peculiar velocities of galaxies. For our default choice of the projected length, $\pi_{\max} = 100 \ h^{-1}$ Mpc, however, the RSD effect is not negligible. We follow the method proposed in Ref. [50]. We employ the linear Kaiser formula [72] to model the RSD effect. We follow Eqs. (51)–(54) in Ref. [50] to model the redshift-space two-point correlation function of galaxies using the Kaiser RSD $\beta$ factor, given by $\beta \equiv (1/b_{\text{eff}})\mathrm{d} \ln D/\mathrm{d} \ln a$, where $b_{\text{eff}}$ is the effective linear bias for a sample of galaxies [see Eq. (19)] and $D$ is the linear growth factor. Then we modify the theoretical template of $w_{\mathrm{p}}$ as

$$w_{\mathrm{p}}(R) \rightarrow f_{\text{RSD}}(R, \pi_{\max})w_{\mathrm{p}}(R; \beta = 0), \quad (30)$$

where $f_{\text{RSD}}(R, \pi_{\max})$ is the correction multiplicative factor to account for the RSD effect. This factor depends on the projected separation $R$ and the projection length $\pi_{\max}$ for an assumed cosmology. For our default choice of $\pi_{\max} = 100 \ h^{-1}$ Mpc, the linear Kaiser factor is a good approximation to model the RSD effect on $w_{\mathrm{p}}$ (see also [26]).

## III. *N*-BODY SIMULATIONS, DARK EMULATOR, AND MOCK CATALOGS OF HSC- AND SDSS-LIKE GALAXIES

In this paper we use two types of mock catalogs of SDSS- and HSC-like galaxies. First, we use high-resolution *N*-body simulations, with periodic boundary conditions, and the halo catalogs to generate the mock catalogs of SDSS galaxies. We then measure $\Delta\Sigma$ and $w_{\mathrm{p}}$ from the mock

catalogs to define the mock signals that we use in cosmology challenges. Second, we also use the mock catalogs of SDSS- and HSC-like galaxies built in the light-cone simulations, including the simulated lensing signals on the HSC-like source galaxies, and use those catalogs to estimate the covariance matrix of the observables. These mocks are the same as those used in our companion paper, Sugiyama *et al.* [26]. In this section, we describe details of *N*-body simulations and the mock catalogs.

### A. *N*-body simulations and Dark Emulator

In this paper we extensively use DARK EMULATOR developed in Nishimichi *et al.* [41], which is a software package enabling fast, accurate computations of halo clustering quantities for a given cosmological model. Here we will briefly review DARK EMULATOR. They constructed an ensemble of cosmological of *N*-body simulations, each of which was performed with $2048^3$ particles for a box with 1 or 2 Gpc/$h$ on a side length, for 101 cosmological models within the flat $w$CDM cosmologies. The $w$CDM cosmology is parametrized by six cosmological parameters, $\mathbf{p} = \{\omega_{\mathrm{b}}, \omega_{\mathrm{c}}, \Omega_{\text{de}}, \ln(10^{10}A_{\mathrm{s}}), n_{\mathrm{s}}, w\}$, where $\omega_{\mathrm{b}}(\equiv \Omega_{\mathrm{b}}h^2)$ and $\omega_{\mathrm{c}}(\equiv \Omega_{\mathrm{c}}h^2)$ are the physical density parameters of baryon and CDM, respectively, $h$ is the Hubble parameter, $\Omega_{\text{de}} \equiv 1 - (\omega_{\mathrm{b}} + \omega_{\mathrm{c}} + \omega_{\nu})/h^2$ is the density parameter of dark energy for a flat-geometry universe, $A_{\mathrm{s}}$ and $n_{\mathrm{s}}$ are the amplitude and tilt parameters of the primordial curvature power spectrum normalized at $k_{\text{pivot}} = 0.05$ Mpc$^{-1}$, and $w$ is the equation of state parameter for dark energy, respectively. For the *N*-body simulations, they included the neutrino effect fixing the neutrino density parameter $\omega_{\nu} \equiv \Omega_{\nu}h^2$ to 0.00064 corresponding to 0.06 eV for the total mass of three neutrino species. They included the effect of massive neutrinos only in the initial linear power spectrum (see [41] for details). To carry out "cosmology challenges" in the following, we employ the fiducial *Planck* cosmology that is characterized by the parameter values in Table I. We use the *N*-body simulation realizations of 1 $h^{-1}$ Gpc box size for the fiducial *Planck* cosmology to construct the mock catalogs for SDSS-like galaxies. The mass of simulation particle for the fiducial *Planck* simulations is $m = 1.02 \times 10^{10} \ h^{-1} \ M_{\odot}$. In the following we use halos with mass greater than $10^{12}h^{-1} \ M_{\odot}$, corresponding to about 100 simulation particles.

For each *N*-body simulation realization (each redshift output) for a given cosmological model, they constructed a catalog of halos using ROCKSTAR [73] that identifies halos and subhalos based on clustering of *N*-body particles in phase space (position and velocity space). The spherical overdensity mass, with respect to the halo center that is defined from the maximum mass density, $M \equiv M_{200 \mathrm{\ m}} = (4\pi/3)R_{200 \mathrm{\ m}}^3 \times (200\bar{\rho}_{\text{m0}})$, is used for definition of halo mass, where $R_{200 \mathrm{\ m}}$ is the spherical halo boundary radius within which the mean mass density is 200 times $\bar{\rho}_{\text{m0}}$.

TABLE I. The value of each cosmological parameter for the fiducial *Planck* cosmology which we use in the following cosmology challenges (we will study whether the method can recover the true values within the credible interval). The column, labeled as "supporting range," gives the supporting range of each parameter in DARK EMULATOR that outputs the halo clustering quantities for a flat-geometry $w$CDM model specified by a set of six cosmological parameters $\{\Omega_{de}, \ln(10^{10}A_s), \omega_b, \omega_c, n_s, w\}$ each of which should be in the supporting range. Here $\sigma_8$ and $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ are derived parameters, and the values in the table are those for the fiducial *Planck* cosmology.

| Parameters | Fiducial value | Supporting range [min,max] |
|---|---|---|
| $\Omega_{de}$ | 0.6844 | [0.54752, 0.82128] |
| $\ln(10^{10}A_s)$ | 3.094 | [2.4752, 3.7128] |
| $\omega_b$ | 0.02225 | [0.0211375, 0.0233625] |
| $\omega_c$ | 0.1198 | [0.10782, 0.13178] |
| $n_s$ | 0.9645 | [0.916275, 1.012725] |
| $w$ | −1 | [−1.2, −0.8] |
| $\sigma_8$ | 0.831 | derived |
| $S_8$ | 0.852 | derived |

By combining the outputs of $N$-body simulations and the halo catalogs at multiple redshifts in the range $z = [0, 1.48]$, they built an emulator, named DARK EMULATOR, which enables fast and accurate computations of the halo mass function, halo-matter cross-correlation, and halo autocorrelation as a function of halo masses, redshift, separations, and cosmological models. It was shown that DARK EMULATOR achieves a sufficient accuracy for these statistical quantities for halos of $10^{13}$ $M_\odot$, which is a typical mass of host halos of SDSS galaxies, compared to the statistical measurement errors of $\Delta\Sigma$ and $w_p$ expected from the HSC and SDSS data, as shown in Fig. 31 of the paper Ref. [41]. In summary, DARK EMULATOR outputs the following

(1) $\frac{dn_h}{dM}(M; z, \mathbf{p})$: the halo mass function for halos in the mass range $[M, M + dM]$ at redshift $z$.

(2) $\xi_{hm}(r; M, z, \mathbf{p})$: the halo-matter cross-correlation function for a sample of halos in the mass range $[M, M + dM]$ at redshift $z$.

(3) $\xi_{hh}(r; M, M', z, \mathbf{p})$: the halo-halo auto-correlation function for two samples of halos with masses $[M, M + dM]$ and $[M', M' + dM']$ at redshift $z$.

for an input set of parameters, halo mass $M$ (and $M'$ for the correlation function of two halo samples), redshift $z$, and cosmological parameters $\mathbf{p}$. In addition, the DARK EMULATOR package outputs auxiliary quantities, based on emulation, such as the linear halo bias (the large-scale limit of the halo bias), the linear mass power spectrum, the linear rms mass fluctuations of halo mass scale $M$ ($\sigma_m^L(M)$), and $\sigma_8$. The supporting range of each of cosmological parameters for DARK EMULATOR is given in Table I. These ranges are sufficiently broad, e.g., to cover the range of cosmological constraints from the current state-of-the-art large-scale structure probe such as the Subaru HSC cosmic shear

results [4,5]. In this paper we will use DARK EMULATOR to perform Markov chain Monte Carlo (MCMC) analyses of cosmological parameters by comparing the model templates of $\Delta\Sigma$ and $w_p$ with the mock signals expected from the SDSS and HSC data.

## B. Mock catalogs for the SDSS-like galaxies for the mock signals of $\Delta\Sigma$ and $w_p$

### 1. Mock catalogs of SDSS LOWZ- and CMASS-like galaxies

We use the $N$-body simulation realizations and the halo catalogs for the *Planck* cosmology to build mock catalogs of galaxies that resemble spectroscopic galaxies in the SDSS-III BOSS DR11 sample [74]. We consider three galaxy samples in three redshift bins: "LOWZ" galaxies in the redshift range $z = [0.15, 0.30]$ and two subsample of "CMASS" galaxies that are obtained from subdivision of CMASS galaxies into two redshift bins, $z = [0.47, 0.55]$ and $z = [0.55, 0.70]$. Here we consider luminosity-limited samples rather than flux-limited samples for these galaxies (see [45] for details; also, see [24] for a similar discussion on the stellar-mass limited sample). The luminosity-limited sample is considered as a nearly volume-limited sample in each redshift bin, and we expect that properties of galaxies do not strongly evolve within the redshift bin, which is desired because we ignore the redshift evolution of clustering observables within the redshift bin for each sample. Table II summarizes characteristics of each galaxy sample.

TABLE II. Specifications of the mock galaxy catalogs that resemble the LOWZ and CMASS galaxies for the spectroscopic SDSS DR11 data. For the CMASS sample we consider two subsamples divided into two redshift ranges. We give the redshift range and the comoving volume for each sample, assuming 8300 deg² for the area coverage. The column denoted as "representative redshift" is a representative redshift of each sample which we assume to represent the clustering observables for each sample (i.e., ignore redshift dependences of the observables within the redshift bin). The lower columns, below double lines, denote the HOD parameters used to build the mock catalogs of each sample from $N$-body simulations for the *Planck* cosmology. Note that $M_{min}$ and $M_1$ are in units of $h^{-1}$ $M_\odot$.

|  | LOWZ | CMASS1 | CMASS2 |
|---|---|---|---|
| Redshift range | [0.15, 0.35] | [0.47, 0.55] | [0.55, 0.70] |
| Representative redshift | 0.251 | 0.484 | 0.617 |
| Volume $[(h^{-1}\,\mathrm{Gpc})^3]$ | 0.67 | 0.81 | 2.00 |
| HOD parameters | | Fiducial values | |
| $\log M_{min}$ | 13.62 | 13.94 | 14.19 |
| $\sigma_{\log M}$ | 0.6915 | 0.7919 | 0.8860 |
| $\kappa$ | 0.51 | 0.60 | 0.066 |
| $\log M_1$ | 14.42 | 14.46 | 14.85 |
| $\alpha$ | 0.9168 | 1.192 | 0.9826 |

To build the mock catalogs for each galaxy sample, we first perform a fitting of the HOD model predictions to the projected correlation function $w_p$ that is actually measured from the SDSS data for each sample, assuming the *Planck* cosmology model, and then estimate the HOD parameters for the best-fit model. Note, however, that we here use the analytical halo model method used in More *et al.* [25], and did not use the weak lensing profile to further constrain the HOD parameters. The fiducial HOD model is the same as that described in Sec. II B. Table II gives the best-fit HOD parameters for each sample.

We use the outputs of $N$-body simulations at $z = 0.251$, 0.484, and 0.617 as representative redshifts of the galaxy samples to build the mock catalogs. We then populate galaxies into halos of each realization for the *Planck* cosmology, using the best-fit HOD parameters in Table II (see also [44,51] for details of the method). For our default mocks, we assume the NFW profile for the radial distribution of satellite galaxies. The default mock is the same as the default HOD model used in the theoretical template (see Sec. II B), so we will use the default mock to perform a sanity check of whether we can recover the true cosmological parameters, i.e., the parameter values of *Planck* cosmology, in the parameter estimation by comparing model templates with the mock signals measured from the default mocks. In the default mocks, we neither include the off-centering effect of central galaxies, an incompleteness effect of central galaxies, nor the redshift-space distortion effect. For other mocks, we include these effects one by one, and then study the impact of each effect on parameter estimation.

### 2. Mock signals of $\Delta\Sigma$ and $w_p$

We generate the mock signals of $\Delta\Sigma$ and $w_p$ for the SDSS-like galaxies, by measuring those clustering observables from each of the mock catalog realizations for the *Planck* cosmology as we described in the preceding section. In doing this, we take an advantage of the periodic boundary conditions in each mock, which allows for a fast computation of the clustering quantities using the FFT algorithm. In addition, the measured clustering quantities are not affected by the window function thanks to the periodic boundary conditions. We here describe details of the measurement method of $\Delta\Sigma$ and $w_p$ from each mock catalog [44,51].

For $\Delta\Sigma$, we first project the matter ($N$-body) particles and galaxies along one axis of $N$-body cubic box assuming that the axis is along the line-of-sight direction, and assign the matter particles and galaxies to the $46,332^2$ two-dimensional grids using the nearest grid point interpolation kernel. We then Fourier transform the matter and galaxy density fields and take a product of the two, $\text{Re}[\tilde{\delta}_g^{2D}(\mathbf{k}_\perp)\tilde{\delta}_m^{2D}(-\mathbf{k}_\perp)]$, where $\mathbf{k}_\perp$ is the two-dimensional wave vector. We perform the inverse 2D Fourier transform to obtain

$$\xi_{gm}^{2D}(\mathbf{R}) = \int \frac{d^2\mathbf{k}_\perp}{(2\pi)^2} e^{i\mathbf{k}_\perp \cdot \mathbf{R}} \text{Re}[\tilde{\delta}_g^{2D}(\mathbf{k}_\perp)\tilde{\delta}_m^{2D}(-\mathbf{k}_\perp)], \quad (31)$$

where $\mathbf{R}$ is the two-dimensional separation vector perpendicular to the projection direction. We perform the azimuthal angle average in an annulus of each radial bin and then obtain the two-dimensional galaxy-matter cross correlation function, $\xi_{gm}^{2D}(R)$.

We compute the projected surface mass density profile around galaxies from $\xi_{gm}^{2D}(R)$, according to Eq. (3), as

$$\Sigma_{gm}(R) \simeq \bar{\rho}_{m0}\xi_{gm}^{2D}(R). \quad (32)$$

The above surface mass density differs from Eq. (3) by a constant additive term (spatially homogeneous term), as this is okay because such a constant term is irrelevant to the weak lensing shear or the excess surface mass density profile. To achieve higher spatial resolution, we measure $\Sigma_{gm}(R)$ using the folding method of the FFT box [75]. Following the method described in Refs. [76,77], we perform multiple measurements in which we fold the box different times (we denote the folding times as $n_{fold}$). We have six measurements of $n_{fold} = \{0, 1, 2, 3, 4, 5\}$ for each catalog, while the grid number $46,332^2$ is kept unchanged at each FFT step. It means that we have measurements that have finer resolutions by up to a factor of $2^5 = 32$. We then combine the $\Sigma_{gm}(R)$ signals obtained by stitching the six measurements between the five boundary scales $\{0.125, 0.25, 0.5, 0.1, 0.2\}$ $h^{-1}$ Mpc. Then we compute the excess surface mass density profile, $\Delta\Sigma(R)$ from $\Sigma(R)$, according to Eq. (1). In each measurement, to achieve better statistics we perform three measurements using the projection along $x$-, $y$-, or $z$-axis directions, and use the average of the three results as the measured signal of $\Delta\Sigma$ for the realization.

For $w_p$, we first assign the mock galaxies to the $1024^3$ three-dimensional FFT grids using the nearest grid point kernel. We then Fourier transform the number density field, and take its square amplitudes $|\tilde{\delta}_g(\mathbf{k})|^2$ for each wave vector $\mathbf{k}$. By performing the inverse Fourier transform, we obtain the estimate of the correlation function at each spatial separation $\mathbf{r} = (R, \pi)$,

$$\xi_{gg}(\mathbf{r}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} |\tilde{\delta}_g(\mathbf{k})|^2. \quad (33)$$

Then we estimate the projected correlation function, $w_p(R)$, from the azimuthal angle average in the circular annulus of each radial bin $R$ and the line-of-sight projection over $\pi = [0, \pi_{max}]$, according to Eq. (7).

We use 19 and 22 independent realizations that are built using the different seeds of the initial conditions for $\Delta\Sigma$ and $w_p$, respectively [41,78]. We then measure the average mock signals of $\Delta\Sigma$ and $w_p$ for each of the galaxy samples (Table II). These correspond to 19 and 22 $(h^{-1}\,\text{Gpc})^3$
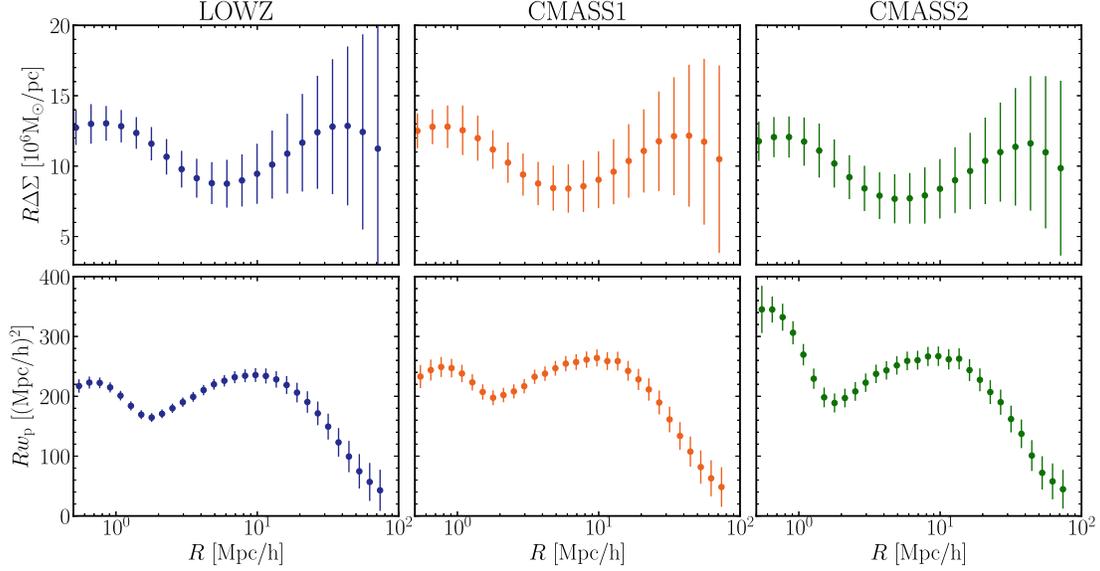
FIG. 4. Mock signals of $\Delta\Sigma$ and $w_p$ for the SDSS LOWZ, CMASS1, and CMASS2 galaxies at the representative redshifts, $z = 0.251$, 0.484, and 0.617, respectively. The signals are computed from the mock catalogs of the galaxies that are constructed by applying the HOD model (Table II) to halo catalogs in high-resolution $N$-body simulations (see text for details). The error bars in each bin denote the statistical errors expected from the SDSS DR11 data (8300 sq. deg.) and the HSC-Y1 data (140 sq. deg.), which are the square root of the diagonal components of each covariance matrix that is estimated from the mock catalogs of HSC and SDSS galaxies in the light cone simulations. Here we multiply $R$ by $\Delta\Sigma(R)$ and $w_p(R)$ for illustrative purpose.

volumes, respectively, that are larger than the volume of any of the three redshift slices in Table II by at least a factor of 11. Recalling that the overlapping area of HSC-Y1 data and SDSS is only about 140 sq. deg., compared to the SDSS area coverage of 8300 sq. deg., the effective volume for the $\Delta\Sigma$ measurement of each sample is smaller than that of $w_p$ by a factor of 60. Hence the simulation volume used for the mock signal of $\Delta\Sigma$ is larger than that of the $\Delta\Sigma$ measurement by at least a factor of 650. Thus by using the mock signals for the larger volumes in cosmology challenges, we can minimize any unwanted bias in estimated parameters due to sample variance. In this way we can evaluate the performance of each method, i.e., the ability to recover the true cosmological parameters, without being affected by the sample variance.

The data points in each panel of Fig. 4 show the mock signals of $\Delta\Sigma$ and $w_p$ for each of the LOWZ, CMASS1, and CMASS2 samples (see Table II) for the *Planck* cosmology. The error bars in each bin are the statistical errors expected for the SDSS and HSC-Y1 surveys as we will explain below. The figure shows that a sufficient number of the realizations of the mock catalogs lead to well-converged, smooth signals in each bin, and the statistical scatters appear to be negligible. This allows us to robustly evaluate the performance of the method for cosmological parameter estimation.

### C. Light cone mock catalogs of HSC- and SDSS-like galaxies for estimating the covariance

As we described in Sec. II A, $\Delta\Sigma$ is independent of source redshift, and depends only on the galaxy-matter

cross correlation at lens redshift. Based on this fact, we construct the best available, accurate mock signal for $\Delta\Sigma$ from the mock catalogs of SDSS-like galaxies as we described above. However, the lensing effects on the same population of source galaxies by foreground structures at different redshifts, from the redshifts of SDSS galaxies, cosmic shear causes statistical scatters in the observed galaxy ellipticities. We need to properly take into account these effects. In this subsection, we describe the mock catalogs of HSC- and SDSS-like galaxies that are built in the light cone simulations, and then use the mock catalogs to model the covariance matrices of $\Delta\Sigma$ and $w_p$.

To construct the mock catalogs in a light cone volume, we use the full-sky, light cone simulations generated in Takahashi *et al.* [79]. The light cone simulation consists of multiple spherical shells with an observer being at the center of the sphere, and each spherical shell contains the lensing fields and the halo distribution, where the lensing fields at the representative redshift of the shell can be used to simulate the lensing distortion effect on a galaxy at the position by foreground structure if the galaxy is located within the shell. The halo distribution in each shell reflects a realization of halos in large-scale structure at the redshift corresponding to the radius of the shell (the distance from an observer to the shell). In this paper we use 108 realizations of the full-sky simulations.

As described in Appendix B in detail, we populate HSC- and SDSS-like galaxies in the full-sky, light cone simulation. For the HSC galaxies, we use the actual HSC shape catalog [43], used for the HSC-Y1 weak lensing

measurements [4], and populate each galaxy into the corresponding shell in the light cone simulation according to its angular position (RA and dec) and photometric redshift (best-fit photo-$z$). Then we simulate the lensing signal on each galaxy using the lensing information of light cone simulation. Thus the mock HSC catalog includes properties of actual data (the angular positions and the distributions of ellipticities and photo-$zs$) as well as the geometry and masks of the HSC footprints. Since the HSC-Y1 data still has a small area coverage (140 sq. deg.), we identify 21 footprints of the HSC-Y1 data in each of the all-sky, light-cone simulation realizations [80,81]. We thus generate 2268 mock catalogs of the HSC data in total.

For the SDSS galaxies, we populate galaxies into halos in the light cone simulation based on the HOD method. We built mock catalogs for each of the three SDSS-like galaxies (LOWZ, CMASS1, and CMASS2) in their corresponding redshift ranges in the assigned survey regions of the SDSS DR11 survey footprints. Given the large survey area of SDSS data (about 8000 sq. deg.), we identify only one SDSS region in each realization of the light cone simulations, and thus build 108 mock catalogs for each of the SDSS galaxies in total.

Each of our mock catalog realizations in the light-cone volume not only contains the angular and redshift (radial) distributions of HSC and SDSS galaxies, but also contains the lensing effects on each source galaxy by the SDSS galaxies and other foreground structures at different redshifts from the SDSS redshift. We then perform measurements of $\Delta\Sigma$ and $w_p$ from each realization using the same analysis pipelines that are used in the actual measurements of the real HSC and SDSS data. Finally we estimate the covariance for the galaxy-galaxy weak lensing $\Delta\Sigma$ and galaxy-galaxy clustering $w_p$ for each of the SDSS LOWZ, CMASS1, and CMASS2 galaxies, from the scatters among the 2268 measurements. The covariance matrices of $\Delta\Sigma$ for the different galaxy samples at different lens redshifts have cross-covariance components because the measurements use the same source galaxies and are affected by weak lensing due to the same foreground structure (cosmic shear) in each light cone simulation realization. Our covariance properly takes into account the cross-covariance.

The error bars in each bin in Fig. 4 are estimated from the covariance matrices we described above. The figure shows that the HSC-Y1 data allows for a significant detection of the lensing signals at each bin. Table III gives the cumulative signal-to-noise (S/N) ratios expected for measurements of $\Delta\Sigma$ and $w_p$ from the HSC-Y1 and SDSS data and the joint measurements. Here, as for our default choice of the range of separation scales, we adopt $3 \leq R/[h^{-1}\,\mathrm{Mpc}] \leq 30$ for $\Delta\Sigma$ and $2 \leq R/[h^{-1}\,\mathrm{Mpc}] \leq 30$ for $w_p$, respectively. The table gives the cumulative S/N ratio integrated over the separation ranges properly taking into account the covariance and the cross-covariance matrices. It is clear that $w_p$ has a grater S/N ratio value than $\Delta\Sigma$ does by

TABLE III. The cumulative signal-to-noise $(S/N)$ ratios of $\Delta\Sigma$, $w_p$ and the joint measurements for the LOWZ, CMASS1, and CMASS2 samples, which are estimated using the mock signals and the covariance matrices. Here we define the "cumulative" $S/N$ over the ranges of $R/[h^{-1}\,\mathrm{Mpc}] = [3, 30]$ and $[2, 30]$ for $\Delta\Sigma$ and $w_p$, respectively, which are our baseline choices of the radial range (see text for details). For the "total" $S/N$ of $\Delta\Sigma$ we take into account the cross-covariances between $\Delta\Sigma$s of different galaxy samples. We assume that the $w_p$ signals for the three samples are independent from each other, and ignore the cross-covariances between $\Delta\Sigma$ and $w_p$.

|  | LOWZ | CMASS1 | CMASS2 | Total |
|---|---|---|---|---|
| $\Delta\Sigma$ | 8.64 | 8.86 | 8.48 | 14.7 |
| $w_p$ | 32.9 | 32.1 | 30.4 | 54.7 |
| joint ($\Delta\Sigma + w_p$) | 34.0 | 33.3 | 31.6 | 56.6 |

a factor of 4, because of the much wider area coverage of SDSS data compared to the HSC-Y1 data by a factor of 60. Nevertheless, we will show later that combining $\Delta\Sigma$ and $w_p$ is crucial to lift parameter degeneracies and obtain useful cosmological constraints.

In the following, we do not include the cosmology dependence of the covariance matrix, motivated by the discussion in [82]. Hence the differences in the performance of different setups/methods that we will show below are purely from the differences in the model or setups.

## IV. METHODOLOGY FOR COSMOLOGY CHALLENGES

The purpose of this paper is to study whether the halo model based method can recover the true cosmological parameters from a hypothetical parameter inference, i.e., comparing the theoretical templates with the mock signals, taking into account the error covariance in the likelihood analysis. Here we describe our methodology to perform cosmology challenges. We employ different setups of the analysis method to quantify the impact of various effects on cosmological parameter estimation, and describe each setup and its purpose.

### A. Parameter estimation method

We assume that the likelihood of the mock signals given the model parameters follows a multivariate Gaussian distribution:

$$\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i,j}[d_i - t_i(\boldsymbol{\theta})](\mathbf{C}^{-1})_{ij}[d_j - t_j(\boldsymbol{\theta})] + \mathrm{const.},$$

$$(34)$$

where $\mathbf{d}$ is the mock data vector, $\mathbf{t}$ is the model vector computed from the theoretical template as a function of the model parameters ($\boldsymbol{\theta}$), $\mathbf{C}^{-1}$ is the inverse of the covariance

matrix, and the summation runs over the indices, $i$, $j$, correspond to the dimension of the data vector. In our baseline analysis we use, as the data vector, $\Delta\Sigma(R)$ given in nine radial bins, logarithmically evenly spaced over $3 \leq R/[h^{-1}\,\mathrm{Mpc}] \leq 30$, and $w_\mathrm{p}(R)$ in 16 radial bins over $2 \leq R/[h^{-1}\,\mathrm{Mpc}] \leq 30$, respectively, for each galaxy sample [83]. Thus we use 75 data points in total $(3 \times (9 + 16) = 75)$. Note that, for our default analysis, we do not include the abundance information of galaxies $(\bar{n}_\mathrm{g})$ in parameter inference. Even if we employ a weak prior on the abundance (e.g., 50% of the number density), the following results remain almost unchanged, but will come back to a question of whether a mild use of the abundance information can improve the parameter estimation.

For the model parameters ($\boldsymbol{\theta}$) in Eq. (34), we consider the two cosmological parameters, $\Omega_\mathrm{de}$ and $\ln(10^{10}A_\mathrm{s})$, and 5 HOD parameters for each of the LOWZ, CMASS1, and CMASS2 samples (Table II). Hence we have 17 parameters $(2 + 3 \times 5 = 17)$ in total. Since the clustering observables are primarily sensitive to the amplitude parameters $\Omega_\mathrm{de}$ and $A_\mathrm{s}$ for the flat $\Lambda$CDM cosmology, we consider only the two parameters, and fix other cosmological parameters to their values of the *Planck* cosmology in parameter inference. Hence the degrees of freedom for the fiducial analysis is $N_\mathrm{dof} = N_\mathrm{d} - N_\mathrm{p} = 75 - 17 = 58$. If we include further nuisance parameters to model the photo-$z$ errors and/or the shear multiplicative bias, we include up to 19 parameters (i.e., 1 or 2 additional parameters). For a given set of the model parameters, we can compute the model vector $\mathbf{t}$ at each radial bin.

We then perform parameter estimation based on the Bayesian inference:

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta}), \qquad (35)$$

where $\mathcal{P}(\boldsymbol{\theta}|\mathbf{d})$ is the posterior distribution of $\boldsymbol{\theta}$ and $\Pi(\boldsymbol{\theta})$ is the prior distribution. Throughout this paper we employ a flat prior on each model parameter as given in Table IV. We checked that the priors of the cosmological parameters are wide enough, and the following results for the posterior distribution are not affected by the prior range.

We draw samples from the posterior distribution of parameters, given the mock signals, with the help of nested sampling as implemented in the publicly available package MULTINEST (Multiodal nested sampler) [84] together with the package MONTE PYTHON [85] to sample the posterior distribution of the parameters. In the following we mainly focus on the posterior distributions of $\Omega_\mathrm{m}$, $\sigma_8$ and $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$. These are derived parameters from the cosmological parameters we use [$\Omega_\mathrm{de}$, $\ln(10^{10}A_\mathrm{s})$], and we employ the definition of $S_8$ following Hikage *et al.* [4] so that the forecast of $S_8$ estimation from our method is compared to the previous results.

In this paper we adopt the mode of the marginalized 1D or 2D posterior distribution to infer the central value(s) of

TABLE IV. Prior range of each model parameter. In cosmology challenges of parameter estimation, we consider two cosmological parameters, $\Omega_\mathrm{de}$, and $\ln(10^{10}A_\mathrm{s})$, and consider 5 HOD parameters for each of the LOWZ, CMASS1, and CMASS2 samples. Hence we have 17 model parameters in total for the baseline method. For an extended method, we also include the nuisance parameters to model the effects of photo-$z$ errors ($\Delta z_\mathrm{ph}$) and the multiplicative shear bias ($m_\gamma$) for which we employ the Gaussian prior with width given in the number in the table.

| Parameters | Prior range [min,max] |
|---|---|
| $\Omega_\mathrm{de}$ | [0.54752, 0.82128] |
| $\ln(10^{10}A_\mathrm{s})$ | [2.4752, 3.7128] |
| $\log M_\mathrm{min}$ | [12.0,14.5] |
| $\sigma^2_{\log M}$ | [0.01,1.0] |
| $\log M_1$ | [12.0,16.0] |
| $\kappa$ | [0.01,3.0] |
| $\alpha$ | [0.5,3.0] |
| $\Delta z_\mathrm{ph}$ | Gauss: 0.04 or 0.1 |
| $m_\gamma$ | Gauss: 0.01 |

the parameter(s), and the highest density interval of the marginalized posterior to infer the credible interval of the parameter(s). We often report the best-fit parameters that correspond to a model at the maximum likelihood in a multidimensional parameter space. As stressed in [26] (see also [86]), a point estimate of parameter is not useful, because it is sensitive to the degree of degeneracies between parameters. For example, even if we consider an ideal case that the input signals are from the model predictions, the central value of a parameter, estimated from the mode of the marginalized posterior distribution, does not necessarily recover the true value as a result of marginalization of the parameters, if the target parameter is highly degenerate with other parameters. Rather a more useful quantity is the credible interval. Hence in the following we will mainly focus on the credible interval, and evaluate each method/setup to study whether the true value of the cosmological parameters are recovered to within the 68% credible interval. We are not interested in an accuracy of recovery of the HOD parameters, and we will not pay much attention to the HOD parameters.

## B. Validation strategy against analysis setups: Scale cuts, parameter degeneracies, observational effects, and RSD

An advantage of the emulator based halo model is that DARK EMULATOR gives an accurate prediction of the halo correlation functions ($\xi_\mathrm{hm}$ and $\xi_\mathrm{hh}$) including all the nonlinear effects down to small scales (nonlinear clustering, nonlinear bias, and the halo exclusion effect) and their dependences on cosmological models within $w$CDM framework. To evaluate the performance of the emulator

TABLE V.   A summary of the analysis setups.

| Setup | Scale cut | Sample parameters | Note |
|---|---|---|---|
| | $[h^{-1}\,\mathrm{Mpc}]$ | | |
| Baseline | (2,3) | $(\Omega_{\mathrm{m}}, \sigma_8)$ + HOD (17 paras.) | Fiducial mock ($R_{\mathrm{max}} = 30\ h^{-1}\,\mathrm{Mpc}$) |
| +RSD | (2,3) | $\cdots$ | Include RSD effect in the fiducial mock |
| Scale cuts | (0.5,0.75) | $\cdots$ | $\cdots$ |
| | (1,1.5) | $\cdots$ | $\cdots$ |
| | (8,12) | $\cdots$ | $\cdots$ |
| $R_{\mathrm{max}} = 70\ h^{-1}\,\mathrm{Mpc}$ | (2,3) | $\cdots$ | Use up to $R_{\mathrm{max}} = 70\ h^{-1}\,\mathrm{Mpc}$ |
| $\Delta\Sigma$ alone | (2,3) | $\cdots$ | $\cdots$ |
| $w_{\mathrm{p}}$ alone | (2,3) | $\cdots$ | $\cdots$ |
| $\Omega_{\mathrm{m}}$-goem. | (2,3) | $\cdots$ | Via $(\Sigma_{\mathrm{cr}}, R)$ |
| Photo-$z$ error ($\Delta z_{\mathrm{ph}}$) | (2,3) | $+\Delta z_{\mathrm{ph}}$ | Only in model |
| Shear-$m$ ($\Delta m_\gamma$) | (2,3) | $+\Delta m_\gamma$ | Only in model |
| $\Omega_{m,\mathrm{geom}} + \Delta z_{\mathrm{ph}} + \Delta m_\gamma$ | (2,3) | $+(\Omega_{m,\mathrm{geom}}, \Delta z_{\mathrm{ph}}, \Delta m_\gamma)$ | Only in model |
| Full HSC | (2,3) | $-$ | Use $0.1 \times \mathrm{Cov}_{\Delta\Sigma}$ mimicking the full HSC data |

based method, we study various setups as summarized in Table V.

One question we want to address in this paper is; which scale cuts for $\Delta\Sigma$ and $w_{\mathrm{p}}$ are adequate in parameter estimation? Since $\Delta\Sigma$ and $w_{\mathrm{p}}$ have higher signal-to-noise ratios at smaller scales, we want to include the information of $\Delta\Sigma$ and $w_{\mathrm{p}}$ down to smaller scales in the nonlinear 1-halo term regime. However, such smaller scales are more affected by nonlinear physics, especially galaxy physics, so it would be difficult to accurately model the clustering signals on very small scales. Including such small-scale information might cause a bias in the estimated cosmological parameters. We should avoid such a failure situation as much as possible. To estimate appropriate scale cuts, we will study the performance of the method adopting different scale cuts of (0.5,0.75), (1,1.5), (2,3), or (8,12) (in units of $h^{-1}\,\mathrm{Mpc}$) for $w_{\mathrm{p}}$ and $\Delta\Sigma$, respectively, where (2,3) is our fiducial choice unless explicitly stated. For all the cases we adopt $R_{\mathrm{max}} = 30\ h^{-1}\,\mathrm{Mpc}$ for the maximum scale up to which we include the information of $\Delta\Sigma$ and $w_{\mathrm{p}}$ for parameter estimation. Thus we do not include the BAO information for all the analyses. To study the impact of the maximum scale, we also study the setup for $R_{\mathrm{max}} = 70\ h^{-1}\,\mathrm{Mpc}$, fixing other parameters to those in the fiducial setups.

For the setups labeled as "$\Delta\Sigma$ alone" and "$w_{\mathrm{p}}$ alone," we study the parameter constraints if using either of $\Delta\Sigma$ or $w_{\mathrm{p}}$ alone. Comparing this result with the baseline method manifests complementarity of $\Delta\Sigma$ and $w_{\mathrm{p}}$ in the cosmological parameter estimation.

To study the impact of the observational effects on parameter estimation, we include the geometrical dependence of $\Omega_{\mathrm{m}}$ and introduce additional parameter to model the

photo-$z$ error and/or multiplicative shear bias, as discussed in Secs. II D 1–II D 3.

For the setup labeled as "RSD," we study the impact of RSD effect. In the theoretical template we model the RSD effect using the linear RSD model. Then we compare the theoretical template with the mock catalog including the full RSD effects, and then assess whether the theoretical model is still applicable to a realistic setup, without any significant bias or degradation in the estimated parameter.

Finally, we show a forecast of how the anticipated full HSC dataset covering 1400 sq. deg., about a factor of 10 larger area than the HSC-Y1 data, can improve the cosmological constraints. To do this forecast, we simply scales the covariance of $\Delta\Sigma$ by the area factor between the HSC-Y1 and full datasets.

### C. Validation strategy against uncertainties in galaxy-halo connection

The HOD model is an empirical prescription of the galaxy-halo connection. Our expectation is that we could recover the underlying cosmological parameters as long as a sufficient number of galaxy-halo connection parameters are introduced and then the effects are marginalized over when estimating cosmological parameters. To assess the "robustness" of our emulator based halo model against uncertainties in galaxy-halo connection, we use a wide variety of mock galaxy catalogs to study whether the baseline method can recover the true cosmological parameters for the different catalogs, as summarized in Table VI.

Figure 5 compares the mock signals of $\Delta\Sigma$ and $w_{\mathrm{p}}$ for different mocks relative to those of the fiducial mock. Here all the mocks, except for the "cent-incomp." and "FOF-halo" mocks, are built using the same HOD as that of the fiducial mocks, but using different ways of

TABLE VI.   A summary of the mock catalogs we use in this paper to assess "robustness" of the baseline method against variations in galaxy-halo connection. All the mock catalogs, except for "`cent-imcomp.`" and "`FOF-halo`" catalogs, have the same HOD in average sense, but leads to modifications in $\Delta\Sigma$ and $w_\mathrm{p}$ in their amplitudes and scale-dependences. The column "satellite gals." denotes a model of the spatial distribution of satellite galaxies in the host halo. In the columns of $\Delta\Sigma$ and $w_\mathrm{p}$, "✓" or "−" denote whether they are modified from the fiducial mock or not, respectively.

| Model | HOD | Satellite gals. | $\Delta\Sigma$ | $w_\mathrm{p}$ | Description |
|---|---|---|---|---|---|
| `fiducial` | fid. | NFW | − | − | Fiducial model |
| `RSD` | fid. | NFW | − | ✓ | Include the RSD effect in $w$ |
| `sat-mod` | fid. | NFW | ✓ | ✓ | Populate satellites irrespectively of centrals |
| `sat-DM` | fid. | DM part. | ✓ | ✓ | Populate satellites according to $N$-body particles |
| `sat-sub` | fid. | Subhalos | ✓ | ✓ | Populate satellites according to subhalos |
| `off-cent1` | fid. | NFW | ✓ | ✓ | All centrals off-centered, with Gaussian profile |
| `off-cent2` | fid. | NFW | ✓ | ✓ | A fraction (0.34) of "off-centered" centrals, assuming Gaussian profile |
| `off-cent3` | fid. | NFW | ✓ | ✓ | Similar to "off-cent1", but with NFW profile |
| `off-cent4` | fid. | NFW | ✓ | ✓ | Similar to "off-cent2", but with NFW profile |
| `cent-incomp.` | $\langle N_\mathrm{c} \rangle$ mod. | NFW | ✓ | ✓ | Include an "incomplete" selection of centrals |
| `FOF-halo` | mod. | FOF halos | ✓ | ✓ | Use FOF halos to populate galaxies |
| `assembly-`$b$`-ext` | fid. | NFW | ✓ | ✓ | Populate galaxies according to concentrations of host halos |
| `assembly-`$b$ | fid. | NFW | ✓ | ✓ | Similar to "assembly-$b$-ext", but introduce scatters |
| `baryon` | fid. | NFW | ✓ | − | Mimic the baryonic effect of Illustris on the halo mass profile |

populating galaxies into individual halos, as described below. The variety in the target observables for a fixed HOD is possible, because the HOD only specifies the average number of galaxies per halo in each mass bin, but there are many different ways to populate galaxies into halos leading to different spatial distributions (clustering properties) of galaxies, even for the same HOD. The different mock catalogs lead to modifications in the
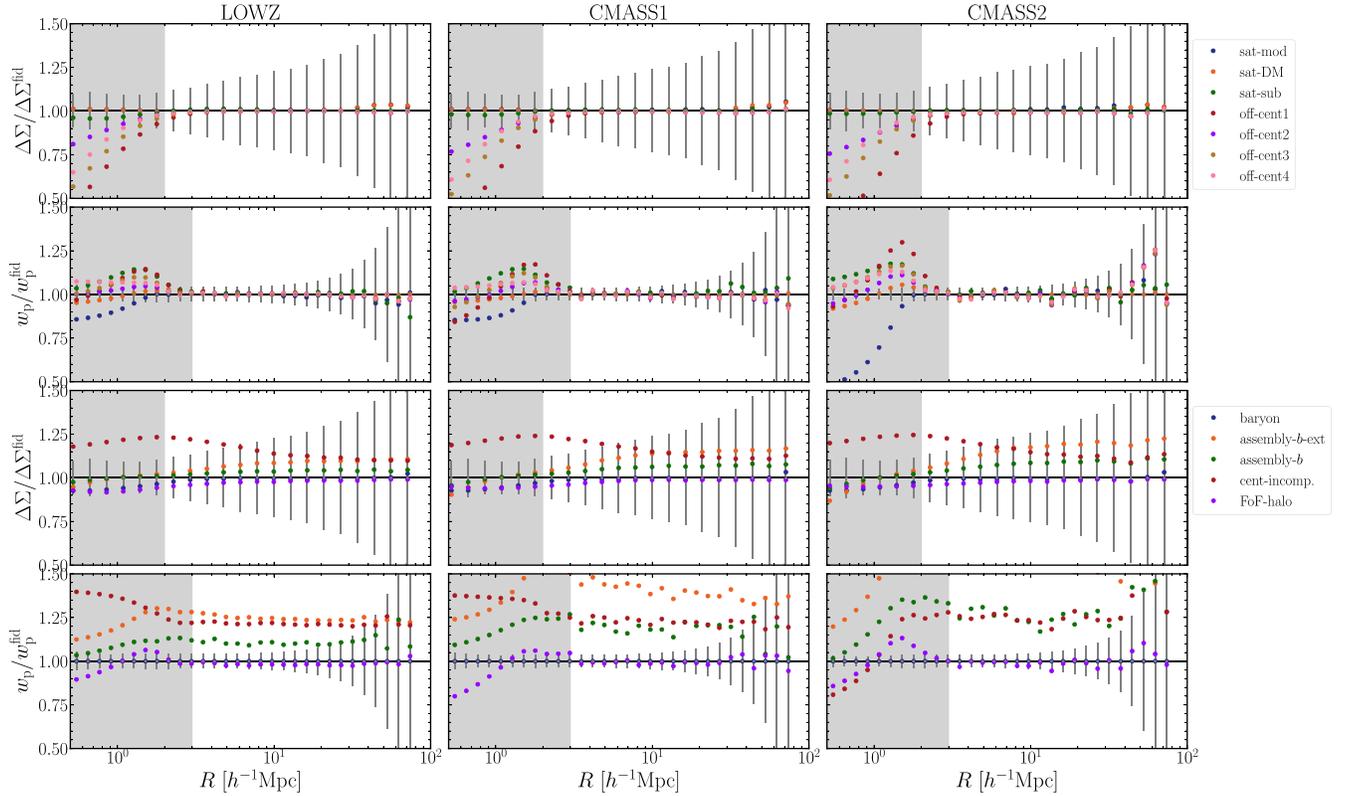


FIG. 5.   The mock signals of $\Delta\Sigma$ and $w_\mathrm{p}$ relative to those for the `fiducial` mock, for each of variants of the mock catalogs in Table VI. The error bars are the same as Fig. 2.

amplitudes and scale-dependences of $\Delta\Sigma$ and $w_\mathrm{p}$ in a complex way, compared to the `fiducial` mock. One of the most important systematic effects is the assembly bias effect, and we also use the mock catalogs, labeled as "`assembly-b-ext`" and "`assembly-b`," to test the performance of our method against the assembly bias effect.

In the following we describe details of each mock. Readers, who are interested in the results, can skip this section and directly go to Sec. V.

### 1. Satellite galaxies

Even if the HOD model is fixed, there are several ways of populating satellites in halos in each simulation realization. To study the impact of variations in the distribution of satellite galaxies, we construct several mocks, for the same HOD as that of the `fiducial` mock.

The "`sat-mod`" mock is a slight modification from the `fiducial` mock. In this mock we populate satellite galaxies in halos irrespective of whether each halo already hosts a central galaxy. In this mock there are halos which host only satellite galaxy(ies) inside, without a central galaxy. Here we assume that the radial distribution of satellite galaxies follows the NFW profile as in the `fiducial` mock.

For the "`sat-DM`" mock, we populate satellite galaxy(ies) in each host halo by randomly assigning each satellite to dark matter particles in the halo, in contrast to the NFW profile.

For the "`sat-subhalo`" mock, we populate satellite galaxy(ies) in each host halo by randomly assigning each satellite to subhalo(s) in the host halo, which are taken from the ROCKSTAR output.

### 2. Off-centering effects of "central" galaxies

For the `fiducial` mocks, we assume that "central" galaxies are located at the center (the highest mass density) of each host halo. However, a central galaxy in a host halo can be "off-centered" as a result of merger or accretion in a hierarchical structure formation [58,59]. To mimic this possible effect, we generate mock catalogs including the off-centering effects of central galaxy in each halo. Note here that we mean, by "central" galaxies, galaxies that are populated into halos according to the central HOD, and the central galaxies can be off-centered from the true halo center. More physically speaking, a galaxy which resides in the most massive subhalo can be considered as a central galaxy, but the galaxy can be off-centered due to the physical effects we discussed above.

We generate four kinds of mock catalogs including the off-centering effects, following the method in Kobayashi *et al.* [51] (see also [21]). The `off-cent1` mock is designed to include the maximum possible amount of off-centering effect, where we assume that all central galaxies are off-centered from the true halo center of each

host halo. We assume that the average radial profile of off-centered galaxies with respect to the halo center follows a Gaussian profile with width $R_\mathrm{off} = 2.2$, i.e., given by $\tilde{p}_\mathrm{off}(k) \propto \exp[-k^2(R_\mathrm{off}r_s)^2/2]$ for the radial profile in Fourier space, where $R_\mathrm{off}$ is a parameter to characterize the typical off-centering radius relative to the scale radius ($r_s$) of the NFW profile. When we populate a central galaxy in each halo, we randomly draw an off-centering radius from the Gaussian profile, and then place the galaxy into the spherical shell of the off-centering radius in the host halo (with randomly choosing the angular position for the azimuthal angles). Then we populate satellite galaxies in the same way as that for the `fiducial` mock.

For the "`off-cent2`" mock, we assume that a fraction of central galaxies are off-centered. Following the implication found in More *et al.* [25,58] for the SDSS galaxies, we assume $q_\mathrm{off} = 0.34$ of the central galaxies are off-centered, while the remaining galaxies ($1 - q_\mathrm{off} = 0.66$) are at the halo center. We then populate satellite galaxies in the same way as that for the `fiducial` mock.

The "`off-cent3`" mock is very similar to the `off-cent1` mock, but we populate the off-centered "central" galaxies into halos assuming that the off-centered galaxies follow the NFW profile of the host halo, similarly to satellite galaxies.

The "`off-cent4`" mock is very similar to the `off-cent2` mock, but we populate the off-centered galaxies assuming the NFW profile as in `off-cent3` mock.

### 3. FOF halos

Dark matter halos are neither uniquely defined objects nor have a clear boundary with the surrounding structure. In this paper we use those identified by the ROCKSTAR algorithm with the spherical-oversensity (SO) masses in each simulation realizations as our default choice (see Sec. III A). For the "`FOF halo`" mock, we use halos that are identified by the friends-of-friends (FOF) method with linking length $b_\mathrm{FOF} = 0.2$, in simulation realizations. The FOF halos do not necessarily have a one-to-one correspondence to the fiducial halos. The mass of each FOF halo is different from the SO mass even if the corresponding halos are identified, because their boundaries are different. Figure 6 compares the mass function of halos measured using the SO halo mass definition or the FOF definition from the same $N$-body simulation realization, at three redshifts corresponding to those for LOWZ, CMASS1, and CMASS2, respectively. The figure shows a sizable difference in the halo mass functions over the range of mass scales we consider.

To generate the "`FOF-halo`" mock we treat the FOF halo mass of individual halos, as the mass argument for the mean HOD functions, and then populate galaxies in FOF halos using the same HOD method as that for the `fiducial` mock.
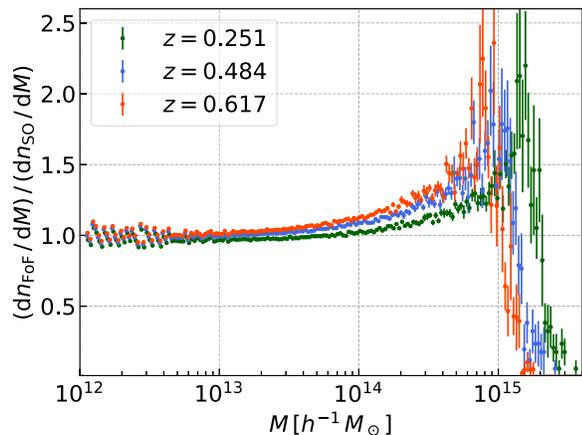
FIG. 6. The ratio of the halo mass function of FOF halos to that of SO halos in each mass bin. The data points show the results at three redshifts $z = 0.251$, $0.484$, and $0.617$, corresponding to the representative redshifts of LOWZ, CMASS1, and CMASS2, respectively. The data point in each bin is the mean of the ratios among 19 realizations of $1(h^{-1}\,\text{Gpc})$-size box simulations, and the error bar is the error on the mean among 19 realizations, which is computed from the scatters among 19 realizations, divided by $\sqrt{19}$.

### 4. Assembly bias effect

The "assembly bias" effect refers to the fact that the clustering amplitudes of galaxies or halos at large scales depends on a secondary parameter other than the halo mass, especially depending on the assembly history of galaxies/halos [87–89]. The assembly bias is one of the most important physical effects causing a violation of the simple halo model picture, which assumes that clustering properties of halos are determined solely by halo mass. To study whether or not cosmology inference based on our method is robust against the assembly bias effect, we use the mocks generated following the method in Kobayashi *et al.* [44]. This is one of the most important tests we address in this paper.

The "assembly-$b$-ext" mock is intended to study the worst case scenario for the assembly bias effect. To make this mock, we first, for each halo, calculate the fraction of mass enclosed within a sphere of 50% of the halo radius $R_{200}$ to the whole halo mass $M_{200}$. We denote this inner mass fraction as $f_{\text{in}}$. We use this quantity as a proxy of the halo concentration, i.e., the higher $f_{\text{in}}$ means the higher concentration. Then we make a ranked list of halos in which we sort the halo by ascending order of $f_{\text{in}}$ in each narrow bin of halo masses. We populate central galaxies, according to the central HOD, into halos from the top of the list (from the lowest-concentration halo) in each mass bin. We then populate satellite galaxies in halos that already host central galaxies using the satellite HOD. For the mock generated in this way, we can have a maximum effect of the assembly bias in the large-scale clustering amplitudes for all the three galaxy samples.

As can be found from Fig. 5, $w_{\text{p}}(R)$ has larger amplitudes at large separations, by up to a factor of 1.6 than that of $w_{\text{p}}$ in the `fiducial` mock, which is quite substantial.

For the mock `assembly-`$b$, we introduce a scatter to $f_{\text{in}}$ of each halo:

$$\log_{10} f_{\text{in}}^{\text{scatter}} = \log_{10} f_{\text{in}} + \epsilon. \qquad (36)$$

We assign a random scatter, $\epsilon$, to each halo drawn from a zero-mean Gaussian distribution with $\sigma$, where $\sigma$ is a parameter to control the amount of the scatter. We adopt $\sigma = 0.1$ to generate the `assembly-`$b$ mock for all the three galaxy samples, which still leads to a significant boost in the clustering amplitudes in $w_{\text{p}}$ by up to a factor of 1.3 (therefore about halved strength compared to the `assembly-`$b$-`ext` mock) than that of $w_{\text{p}}$ in the `fiducial` mock. This might be more realistic for host halos of the SDSS galaxies ($\sim 10^{13} h^{-1}\,M_{\odot}$), although the assembly bias effect has not been detected at a high significance from the real data. Note that the mean HOD is not modified from the `fiducial` mocks with these procedures.

### 5. Baryonic effect

The baryonic effects inherent in galaxy formation/evolution, which we are missing in our $N$-body simulations, are another important physical systematic effect, and we need to quantify their impact on the parameter inference in our cosmology challenges. Although the baryonic effects are still difficult to accurately model from the first principles, one should keep in mind some conservation properties in the distribution of galaxies. First, massive galaxies like the SDSS LOWZ/CMASS galaxies are likely to form at the same peaks of primordial density fluctuations even in the presence of baryonic physics. Hence the distribution of massive galaxies relative to the total matter distribution, i.e., the bias function, is not largely changed by baryonic physics [19,90]. On the other hand, the baryonic physics causes a redistribution of matter around each galaxy, e.g., due to various effects such as dissipative contraction and supernova/AGN feedbacks. For this reason, the radial profile of matter distribution around a galaxy would likely be modified.

We follow the method in Schneider and Teyssier [48] (see also [49]) to include baryonic effects to the mock signals. The notable feature of this model is that the model explicitly imposes the mass conservation around halos, and models the baryonic effects as a redistribution of the surrounding matter around each halo (more exactly the halo profile). We tuned the model parameters so that the model prediction reproduces the baryonic effect on the lensing profile, $\Delta\Sigma$, in the original ILLUSTRIS hydrosimulation [18] that implemented too large baryonic effect than implied by observations, as explicitly demonstrated in Fig. 7. Hence the `baryon` mocks are considered as a worst case of the baryonic effect on the weak lensing profile.
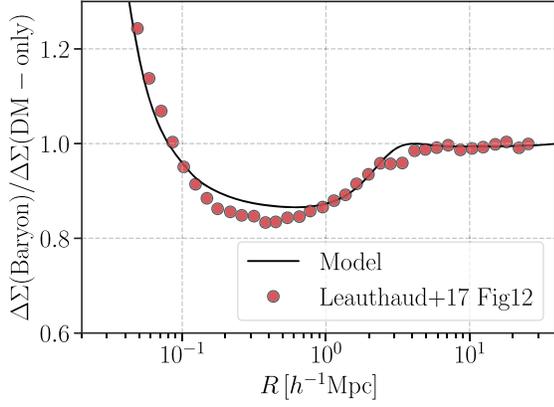
FIG. 7.   We use the method in Schneider and Teyssier [48] to model the baryonic effect on the galaxy-galaxy weak lensing profile, $\Delta\Sigma$, where the mass conservation around halos hosting galaxies is explicitly imposed. We tune the model parameters to reproduce the baryonic effect that is seen in the ILLUSTRIS hydrosimulation [18]. Shown is the ratio of $\Delta\Sigma$ for the matched host halos of SDSS-like galaxies in the simulations with and without the baryonic effect, which is taken from Fig. 12 in Leauthaud *et al.* [91]. Our method nicely capture the baryonic effect, but we note that this would be a worst case of the baryonic effect, because the ILLUSTRIS (not ILLUSTRISTNG) employed the too large baryonic effect (especially AGN feedback).

For $w_{\rm p}$, we use the mock signal for the `fiducial` mocks; that is, we do not include a possible effect of the baryonic physics on the distribution of galaxies in the host halo.

### D. Cross-correlation coefficients in the mocks

All the mock catalogs we use in this paper have a property that the cross-correlation coefficient, $r_{\rm cc}(r) \equiv \xi_{\rm gm}/[\xi_{\rm gg}\xi_{\rm mm}]^{1/2} \simeq 1$ at the limit of large scales, greater than the size of massive halos, say at $\gtrsim 10\ h^{-1}$ Mpc, as shown in Fig. 8. This property is expected if galaxy physics is confined to local, small scales, and because the clustering amplitudes at larger scales than the scale of galaxy physics are governed by gravity alone (and properties of primordial

fluctuations) and then behave as a linear biasing relation to the underlying matter distribution at the large scales. Encouragingly this is confirmed recently by using the hydrodynamical simulation ILLUSTRISTNG in Ref. [62], where various galaxy samples, which are selected based on host halos and the various environment parameters, display $r_{\rm cc} \simeq 1$ at $r \gtrsim 10\ h^{-1}$ Mpc. The DARK EMULATOR outputs also predict $r \equiv \xi_{\rm hm}/[\xi_{\rm hh}\xi_{\rm mm}]^{1/2} \simeq 1$ for halo correlation functions at scales greater than the size of halos. However, even for the scales where $r_{\rm cc} \simeq 1$, this does not mean that the linear theory serves as an accurate theoretical template [26]. It is important to include the nonlinear clustering, the nonlinear halo bias and the halo exclusion effect in the theoretical template.

We also note that $r_{\rm cc}$ can be greater than unity; $r_{\rm cc} \geq 1$, which occurs in the 1-halo term regime where the sub-Poisson nature $\langle N_{\rm g}(N_{\rm g} - 1)\rangle \neq \langle N_{\rm g}\rangle^2$ becomes important due to a finite number statistics of galaxies in the same host halo [25,34,37].

## V. RESULTS

In this section we show the main results of this paper, which are assessment and validation of the emulator based halo model method for cosmology inference. Here we mean by "validation" whether the method can recover the true cosmological parameters, $\Omega_{\rm m0}, \sigma_8$, and $S_8$, to within 68% credible intervals, after marginalization over the HOD parameters (galaxy-halo connection parameters). Note that we will not focus on the accuracy of recovering the HOD parameters.

### A. Validation of the baseline method, complementarity of $\Delta\Sigma$ and $w_{\rm p}$, and tests with different scale cuts

First we perform a sanity check; we study whether our baseline method (see Table V) can recover the true cosmological parameters when comparing the theoretical templates to the mock signals measured from the `fiducial` mock that is based on the same HOD model used in
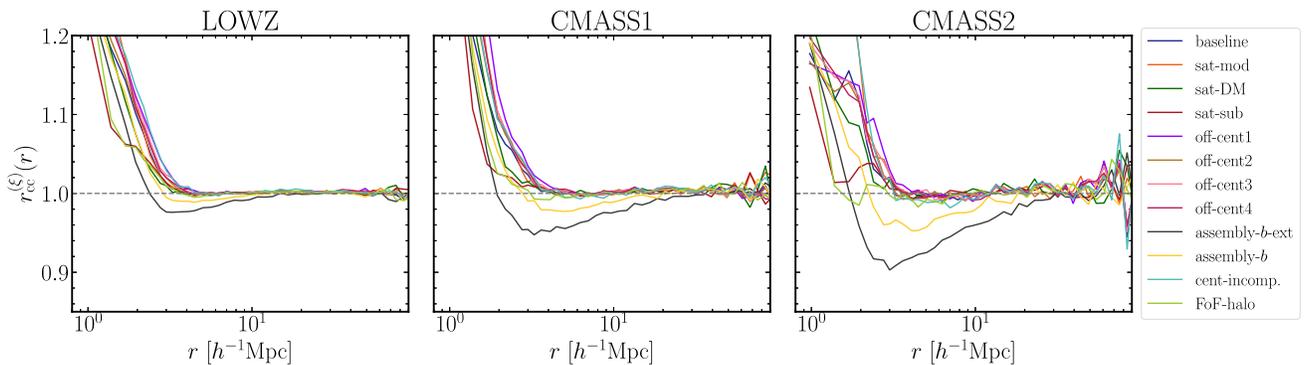


FIG. 8.   Cross correlation coefficients, defined as $r_{\rm cc}^{(\xi)}(r) \equiv \xi_{\rm gm}(r)/\sqrt{\xi_{\rm gg}(r)\xi_{\rm mm}(r)}$, for different mock catalogs of SDSS-like galaxies we use in this paper.

the theoretical template. Different model parameters affect the observables in a complex way, so it is not obvious whether the baseline method can recover the true cosmological parameters, after projecting the posterior distribution in a multidimensional parameter space onto a subspace including the cosmological parameters. In fact, as discussed in Refs. [26,86], the modes of the marginalized posterior distribution of cosmological parameters could be biased from the true values, if the parameters suffer from severe degeneracies.

Figure 9 shows the results. Here we employ $R_{\rm cut} = 2$ and $3h^{-1}$ Mpc for the scale cuts for $w_{\rm p}$ and $\Delta\Sigma$ above which we
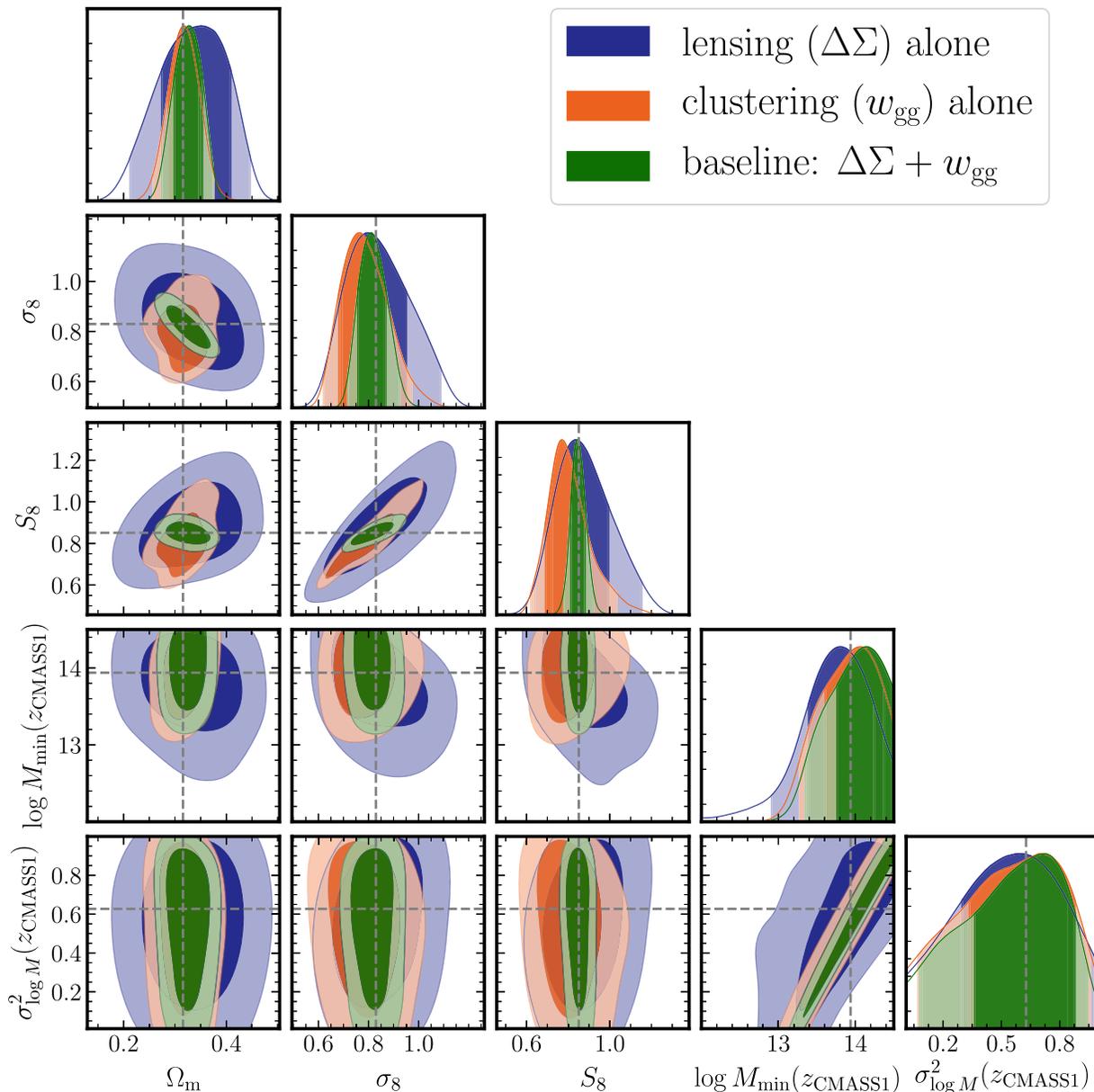


FIG. 9.   Marginalized posterior distribution in each 2D subspace of the parameters, obtained from the projected correlation function information alone ($w_{\rm p}$; orange-color contours), the lensing information alone ($\Delta\Sigma$; blue) and the joint constraints (green), respectively. The inner and outer contours show the 68% and 95% credible regions, respectively. We adopt the baseline setup for which we employ $R_{\rm cut} = 2$ and $3~h^{-1}$ Mpc for the scales cuts of $w_{\rm p}$ and $\Delta\Sigma$, respectively. For the mock signals, we use the signals measured from the `fiducial` mocks which are generated using the same HOD model as those in the theoretical templates. Here we include 17 model parameters in the parameter inference: two cosmological parameters ($\Omega_{\rm m}$ and $\sigma_8$) plus five HOD parameters for each of the three galaxy samples (LOWZ, CMASS1, and CMASS2). Here we show, as an example, the results for the central HOD parameters for the CMASS1 sample at $z = 0.484$, i.e., $M_{\rm min}$ and $\sigma_{\log M}$. The vertical and horizontal dashed lines denote the input parameter value used in the mock catalog.

include the clustering and lensing information in parameter inference (see Table V). The figure nicely shows that the lensing and clustering information are complementary to each other, and combining the two lifts the parameter degeneracies. Thus Fig. 9 gives a validation of the baseline method; the baseline method can recover the cosmological parameters if properties of the galaxy clustering for SDSS-like galaxies are close to the fiducial HOD model we employed. The baseline method achieves a precision of $\sigma(S_8) \simeq 0.035$ for the SDSS and HSC-Y1 data. In the following we show mainly the results for the joint probes combining the information of $\Delta\Sigma$ and $w_p$. For completeness of our discussion, in Fig. 22 we show the posterior distribution of the parameters including all the HOD parameters for the CMASS1-like galaxy sample.

What is the impact of HOD parameters on cosmological parameter estimation? To answer this question, in Fig. 10 we show the results when fixing the HOD parameters to their fiducial values. It is clear that the HOD parameters cause significant degradations in the cosmological parameters; for example, the marginalized error of $\Omega_m$ is enlarged by a factor of 3 when including the HOD parameters. However, this is a price one has to pay to obtain robust constraints on cosmological parameters when including the small-scale information. If one uses a more aggressive method, e.g., by using a less flexible model of galaxy-halo connection (such as fixing some HOD parameters), then one could suffer from severe biases in cosmological parameters.
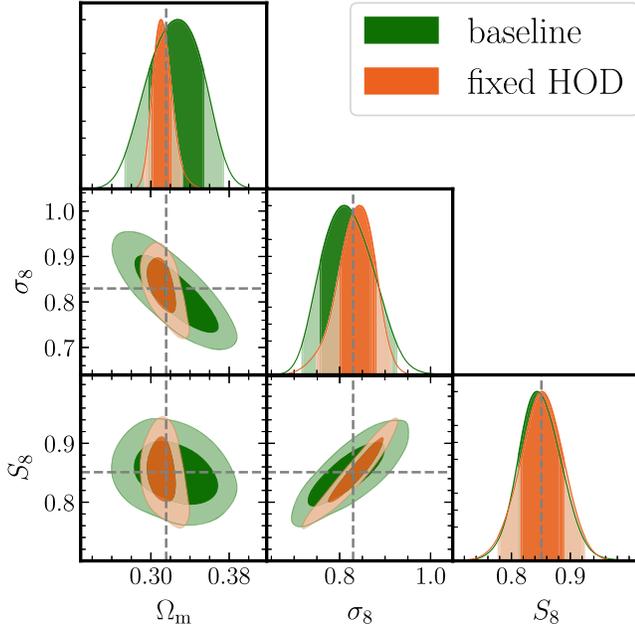
Our results might also be compared to a more conservative approach, e.g., a method using only the clustering observables at large scales, i.e., not including the small-scale information. In Fig. 11 we compare the results from the halo model method, studied in this paper, with those from the PT based method studied in Sugiyama *et al.* [26]. They employed the "minimal"-bias model using the fully nonlinear matter power spectrum and the linear bias parameter $(b_1)$: $\xi_{gm} = b_1 \xi_{mm}^{NL}$ and $\xi_{gg} = b_1^2 \xi_{mm}^{NL}$, where the halofit model is used to model the nonlinear $\xi_{mm}^{NL}$. It was shown that, as long as the conservative scale cuts $R_{cut} = 12$ and 8 $h^{-1}$ Mpc for $\Delta\Sigma$ and $w_p$ are employed and the bias parameter is treated as a free parameter, then the minimal bias method passes all the validation tests against a variety of the mock catalogs including the assembly bias mock. Hence, the results for the minimal bias method can be considered as conservative, yet robust parameter constraints that can be extracted from the SDSS and HSC-Y1 data. The figure clearly shows that including the small-scale information and the halo bias information can improve cosmological constraints compared to the conservative method.
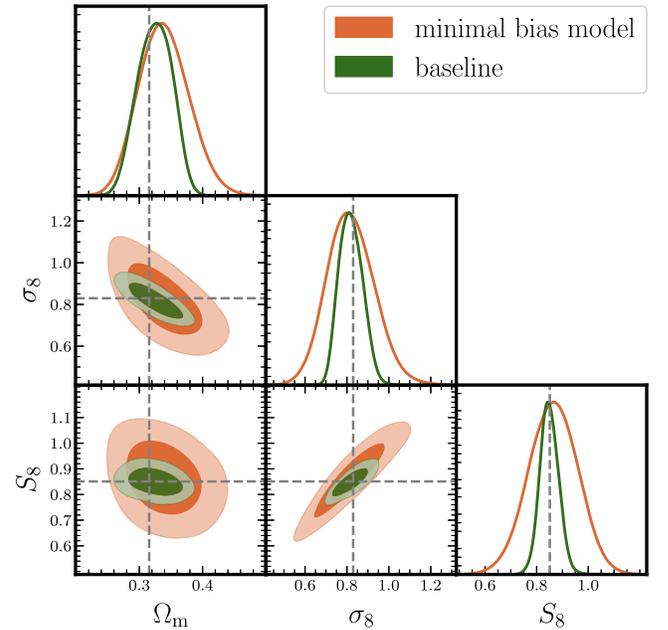


FIG. 11. Similar to Fig. 9, but the figure compares the marginalized posterior distributions obtained from the HOD based method, studied in this paper, and the perturbation theory (PT) inspired method in Sugiyama *et al.* [26]. Here we apply both the methods to the same `fiducial` mock signals. For the PT method, we model the matter-galaxy cross-correlation ($\xi_{gm}$) and the galaxy autocorrelation ($\xi_{gg}$) by the nonlinear matter autocorrelation multiplied by a linear bias parameter: $\xi_{gm} = b_1 \xi_{mm}^{NL}$ and $\xi_{gg} = b_1^2 \xi_{mm}^{NL}$. We then treat $b_1$ as a free parameter in the parameter inference, and employ the scale cuts of $R_{cut} = 12$ and 8 $h^{-1}$ Mpc for $\Delta\Sigma$ and $w_p$, respectively, compared to $R_{cut} = 3$ and 2 $h^{-1}$ Mpc for the halo model based method. The green contours are the same as those in Fig. 9.
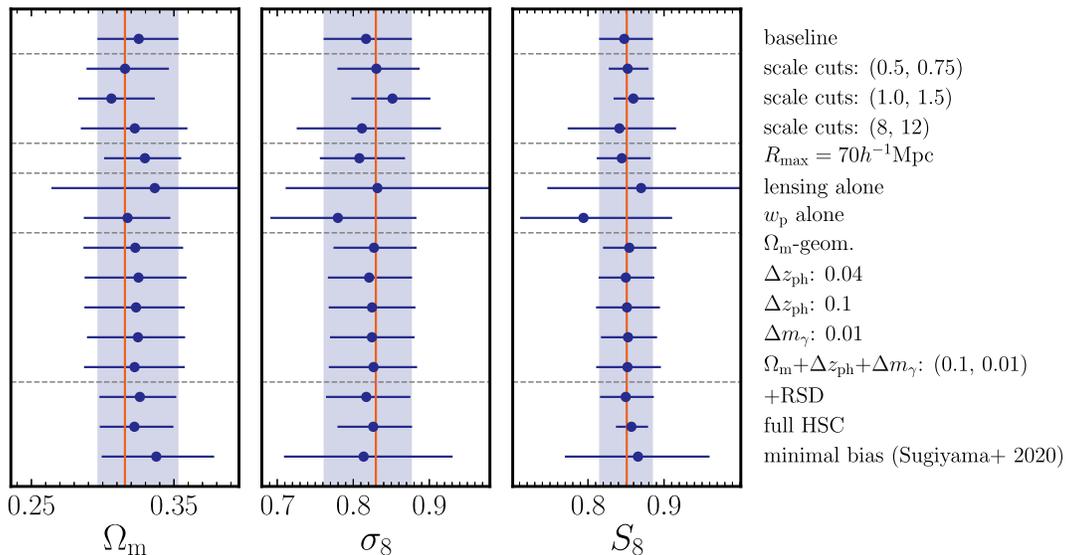


FIG. 10. Similar to the previous figure, but shown are the results when fixing the HOD parameters in the model templates to their fiducial (input) values. The green contours (labeled as "baseline") are the same as those in Fig. 9.

FIG. 12. Summary of the estimation of each cosmological parameter, $\Omega_m$, $\sigma_8$, or $S_8 (\equiv \sigma_8 \Omega_m^{0.5})$, for the different setups in Table V. The blue dot in each row denotes the mode of the marginalized posterior distribution of each parameter, and the error bar denotes the 68% credible interval, which is computed from the highest density interval of the marginalized 1D posterior distribution. The vertical red line denotes the true value used in the mock catalog, and the shaded region denotes the 68% credible interval for the baseline setup for comparison.

In particular, the halo model based method gives a factor of 2 or 3 improvement in the marginalized error of $\sigma_8$ or $S_8$, respectively, compared to that for the minimal bias method. Thus the halo model method has a potential to obtain the improved cosmological constraints, if the model is flexible enough to describe variations in galaxy clustering at small scales down to a few Mpc, which we will test later.

In Fig. 12 we give a summary of the performance for different setups. The second to the fourth row, lying between the horizontal dashed lines, show the results when using the different scale cuts, $(R_{cut}^{\Delta\Sigma}, R_{cut}^{w_p}) = (0.5, 0.75), (1.0, 1.5)$, or $(8, 12) h^{-1}$ Mpc, instead of $(2,3)$ as our fiducial choice. For the `fiducial` mocks, these smaller scale cuts apparently recover the cosmological parameters; the true value of each parameter is within the 68% credible interval. However, a closer look also reveals that the size of the credible interval is not significantly improved by including the smaller information. We also checked that, when applying the smaller scale cuts to other mock catalogs rather than the `fiducial` mocks, those lead to a larger bias in the cosmological parameters compared to the baseline method. Hence, with these results, we conclude that the fiducial scale cuts of $(2, 3)$ $h^{-1}$ Mpc are reasonable. On the other hand, the row labeled as "$R_{max} = 70$ $h^{-1}$ Mpc" shows the results when including the information of $\Delta\Sigma$ and $w_p$ up to $R_{max} \simeq 70 h^{-1}$ Mpc instead of our default choice $R_{max} = 30 h^{-1}$ Mpc. Note that this maximum scale is still below the BAO scales, and we do not include the BAO information. It is clear that the larger-scale information does not improve the cosmological parameter estimation, and $30 h^{-1}$ Mpc seems sufficient for the signal-to-noise level expected for the HSC-SDSS analysis.

### B. Comparison with analytic halo model

In Fig. 13, we compare the results for our baseline method with those obtained by comparing the model predictions of analytical halo model [92] in More *et al.*
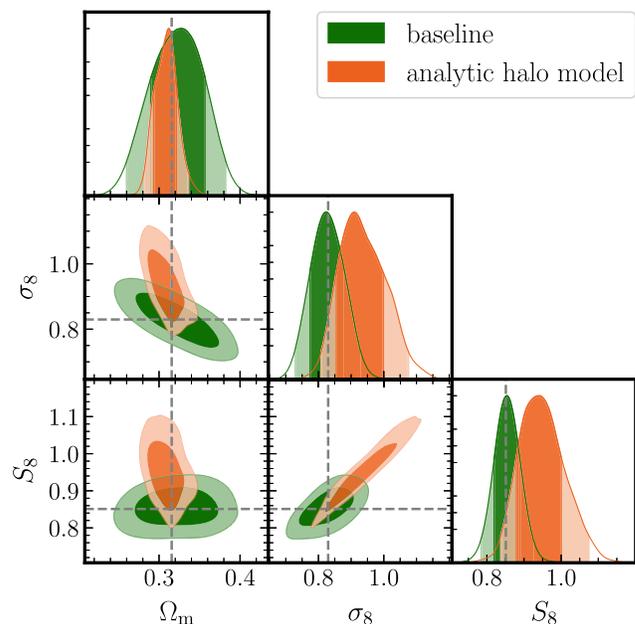


FIG. 13. Similar to Fig. 9, but the plot shows the results for the baseline method with those obtained by using the analytic halo model in More *et al.* [25]. Here we used the same scale cuts, $R_{cut} = 2$ and $3 h^{-1}$ Mpc, for $w_p$ and $\Delta\Sigma$, respectively, for both methods.

[25] with the mock signals from the `fiducial` mocks. The analytical halo model was constructed by calibrating the standard halo model [34] with $N$-body simulation results available at that time, e.g., to include the halo exclusion effect and the nonlinear halo bias effect (see Ref. [50] for details). For example, the analytical halo model uses the fitting formula of halo bias developed in Refs. [93,94], but several recent studies point out an inaccuracy in the fitting formula, up to 5% in the bias amplitude (see the bottom panel of Fig. 22 in Ref. [41]; see also Refs. [95–97]). For the analytical halo model, we did not include the off-centering effect or incompleteness effect as in our baseline method.

The figure shows sizable differences in all the cosmological parameters between our emulator based method and the analytical halo model. The 1D posterior distribution shows that the analytical halo model cannot recover the true values of $\sigma_8$ and $S_8$ to within the 68% credible interval. In addition, the degeneracy directions in the projected two parameter subspace are quite different. Figuring out the cause of the difference is beyond the scope of this paper, but here at least we would like to stress that the emulator-based halo model displays a better performance.

### C. The impacts of observational effects: The geometrical correction, photo-$z$ errors, shear multiplicative bias, and RSD

As discussed in Sec. II D, various observational effects could affect the cosmological inference from the measured $\Delta\Sigma$ and $w_p$. For the results discussed up to the preceding subsections, we ignored these effects, and in this section we study the impact of these effects.

The rows with labels starting from "$\Omega_m - $ geom." to "+RSD" in Fig. 12 show the results when including each or all of these observational effects. First let us discuss the results for "+RSD." For this test we use the mock catalogs where we use the fiducial HOD to populate galaxies into the simulation realizations and include the RSD effects due to the peculiar velocities of individual galaxies [44,51]. Then we perform the cosmology inference by comparing the theoretical templates, including the linear Kaiser effect (Sec. II D 4), with the mock signals. The figure shows that the linear Kaiser model can properly take into account the RSD effect for our fiducial projection length ($\pi_{\max} = 100 \ h^{-1}$ Mpc), and there is no degradation in the parameter estimation. This is similar to the result in Sugiyama *et al.* [26]. Although the RSD effect itself carries the dependence on $\Omega_m$ (more exactly via the growth rate), it leads to only slight improvement in $\Omega_m$.

Now we discuss the results for "$\Omega_m - $ geom.," which refers to the fact that a reference cosmology needs to be assumed to measure $\Delta\Sigma(R)$ and $w_p(R)$ from direct observables, and the assumed cosmology generally differs from the underlying true cosmology. Exactly speaking, for a flat-geometry $\Lambda$CDM model, we have to "remeasure" $\Delta\Sigma(R)$

and $w_p(R)$ every time $\Omega_m$ is varied in parameter estimation. This might be time consuming, and indeed More [64] showed that this conversion can be safely done by a multiplicative factor taking into account the geometrical dependence, which we employ in this paper. Figure 12 shows that including the $\Omega_m$ geometrical dependence does not cause a bias in the cosmological parameters. However, a closer look shows that including this dependence slightly enlarges the credible interval. This happens as follows. If we assume a slightly larger value of $\Omega_m$, then it leads to the smaller amplitude in the model prediction of $w_p$ (recall that the current constraint is mainly from the $w_p$ information). However, when assuming such a model with increased $\Omega_m$ in the measurements, it leads to the smaller amplitudes in the measured $w_p$ as shown in Fig. 2. Thus the dependences of $\Omega_m$ are compensated to some extent in the model and the measurement. This is the reason that the credible interval of $\Omega_m$ is slightly degraded.

Next we discuss the impacts of photo-$z$ errors and the multiplicative shear bias, which are among the most important systematic errors in the weak lensing measurements. As discussed in Secs. II D 2 and II D 3, we introduce nuisance parameters, denoted as $\Delta z_{ph}$ and $\Delta m_\gamma$, to model these effects. To do this we employ the parameters to model the systematic effects that could be present for the actual HSC data. For a sample of source galaxies, we assume that source galaxies are selected based on their photo-$z$s satisfying a conservative cut that the photo-$z$ posterior distribution of individual galaxies satisfies $\int_{0.75}^{\infty} dz \, p(z) \geq 0.99$ [98,99], where the lower cutoff of the integration, $z = 0.75$, is well above the redshifts of CMASS galaxies (the maximum redshift cut of CMASS galaxies is 0.7 as shown in Table II). After using the reweighting method in [4] to infer the intrinsic redshift distribution of source galaxies, we compute the average of the lensing critical surface density over the source distribution, $\langle \Sigma_{cr}^{-1} \rangle$, which is used in the $\Delta\Sigma$ measurement for each sample of lensing galaxies (LOWZ, CMASS1, and CMASS2). Then, by shifting the posterior distribution of all the source galaxies by the same amount, $\Delta z_{ph}$, we repeat the same calculation of $\langle \Sigma_{cr}^{-1} \rangle$ to estimate a shift in the $\Delta\Sigma$ signal. After computing these quantities for multiple values of $\Delta z_{ph}$, we include $\Delta z_{ph}$ to compute the shift in $\Delta\Sigma$ from an interpolation of the precomputed values of $\langle \Sigma_{cr}^{-1} \rangle$. In this paper we study the impact of photo-$z$ errors on parameter estimation by adopting the Gaussian prior with width $\sigma(\Delta z_{ph}) = 0.04$ or 0.1. Here $\sigma(z_{ph}) = 0.04$ corresponds to the quoted errors of photo-$z$ estimation for the source galaxies, estimated using the same method in [4]. We also adopt a very conservative prior of $\sigma(\Delta z_{ph}) = 0.1$, which is about a factor of 2.5 larger than the quoted errors. Figure 12 shows that the cosmological parameters remain almost unchanged by the photo-$z$ errors. Thus, as claimed in Oguri and Takada [21], a self-calibration of the photo-$z$ errors

seems to work very well, by using the single sample of source galaxies and utilizing the dependence of lensing efficiency on lens redshifts.

For a multiplicative shear bias, $\Delta m_\gamma$, we employ the Gaussian prior $\sigma_{\Delta m_\gamma} = 0.01$ as recommended based on dedicated image simulations of HSC galaxies in Mandelbaum *et al.* [43,100]. Figure 12 shows that the multiplicative shear bias does not either affect the cosmological parameters. Thus, the method of Oguri and Takada [21] allows us to self-calibrate both the photo-$z$ errors and the multiplicative bias errors. The row denoted as "$\Omega_m + \Delta z_{ph} + \Delta m_\gamma$" shows the results including all these three effects. The results are not so different from the baseline method ignoring these effects. Hence we conclude that these observational effects are not a severe source of systematic errors causing a sizable bias in the cosmological parameters.

When including these observational effects, the baseline method can achieve a precision of $\sigma(S_8) \simeq 0.042$.

### D. Tests and validations against uncertainties in galaxy-halo connection

In this section, to assess the performance of the halo model method against uncertainties in the galaxy-halo connection, we perform the parameter inference against various mock catalogs (Table VI). Note that we here employ the fiducial HOD model (five parameters model), which does not include the off-centering effect, the incompleteness effect, the baryonic effect, nor the assembly bias effect. Here we address whether the fiducial halo model can recover the true cosmological parameters to within the credible intervals after marginalizing over the HOD parameters. Even if we use a weak prior on the number density of galaxies (the abundance), e.g., 50% or 5% of the number density, the following results remain almost unchanged.

Figure 14 gives a summary of the results. The figure shows that, except for the assembly bias mocks (and the extreme off-centering mock, off-cent1, for $\Omega_{m0}$ and $\sigma_8$), the halo model method with the baseline setup recovers the input cosmological parameters to within the 68% credible interval. We should also note that $S_8$ is better recovered compared to $\Omega_{m0}$ or $\sigma_8$. This is an encouraging result, because $S_8$ is close to the primary parameter combination that determines the amplitudes of $w_p$ and $\Delta\Sigma$. We below discuss the results for some variants of the mocks in more detail.

Figures 15 and 16 show the marginalized distributions of the cosmological parameters for the off-cent, FOF-halo, and cent-incomp. mocks (see Table VI). For most cases except for the off-cent1 mock, where all the central galaxies are off-centered, the halo model method recovers the true cosmological parameters within 68% credible interval. As shown in Fig. 5, the off-cent1
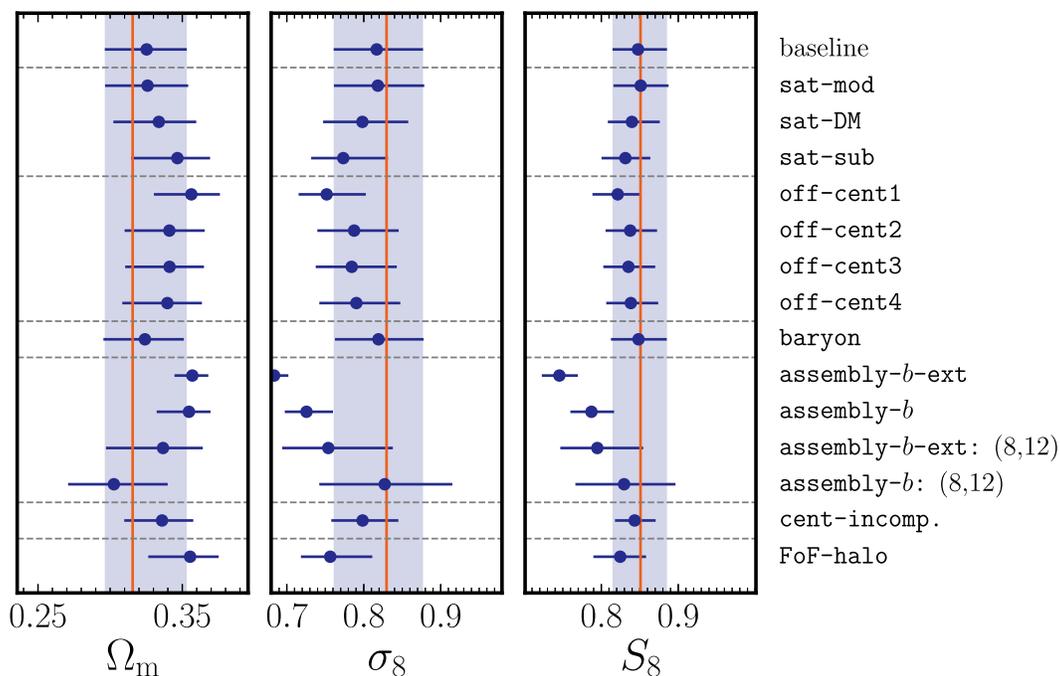


FIG. 14. A summary of the performance of the baseline method against different mock catalogs. The circle shows the mode of the marginalized posterior distribution of each cosmological parameter, and the error bar denotes the 68% credible interval. The vertical solid line denotes the true value of each parameter. The shaded region denotes the 68% credible interval for the fiducial mock, for comparison. The columns labeled as "assembly-*b*: (8,12)" and "assembly-*b*-ext: (8,12)" show the results when the larger scale cuts of (8,12) $h^{-1}$ Mpc are employed for $w_p$ and $\Delta\Sigma$.
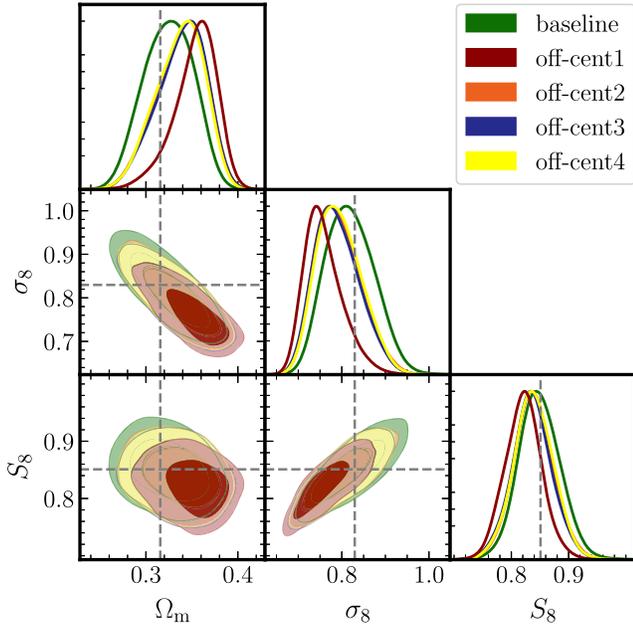
FIG. 15. Marginalized posterior distributions of the cosmological parameters when applying the halo model method to the mock catalog where the off-centering effect of central galaxies is included (Table VI). For comparison, we show the result for the `fiducial` mock, the same as the green contours in Fig. 9.

mock gives smaller amplitudes in $\Delta\Sigma$ at scales around the scale cut ($R_{\mathrm{cut}} = 3 \ h^{-1}$ Mpc), but does not change the $w_{\mathrm{p}}$ amplitude at scales greater than the scale cut ($R = 2 \ h^{-1}$ Mpc). The smaller lensing amplitude leads to an underestimation in the average mass of host halos.
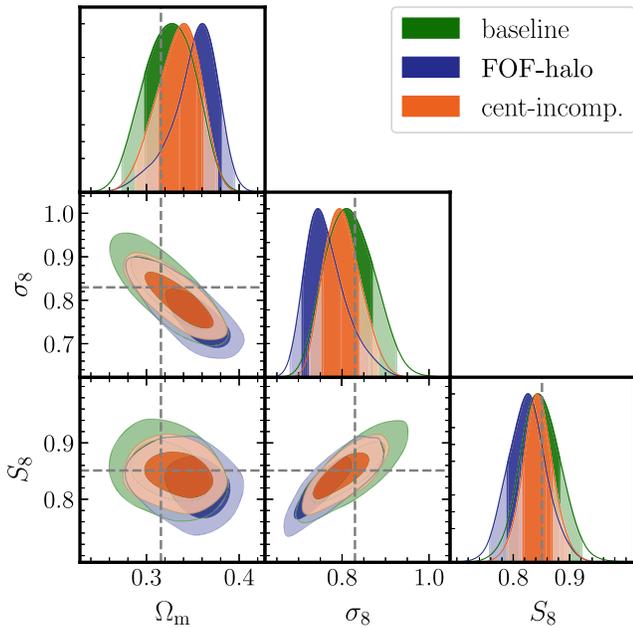


FIG. 16. Similar to the previous figure, but shown is the result for the `FOF` and `cent-incomp.` mocks.

Since smaller-mass halos give a smaller bias amplitude for a fixed cosmology, this would lead to a lower $\sigma_8$ to reproduce the amplitude of $w_{\mathrm{p}}$ in the mock signal (see Fig. 2). This explains a negative bias in $\sigma_8$ for the `off-cent1` mock. We confirmed that, if we employ the larger scale cut for $\Delta\Sigma$, then the bias in $\sigma_8$ is mitigated. For the `FOF-halo` mock, our method nicely recovers $S_8$, although we find sizable biases in $\Omega_{\mathrm{m0}}$ and $\sigma_8$.

In Fig. 17, we show the results for the `baryon`, `assembly-b`, and `assembly-b-ext` mocks. The figure shows that the assembly bias mocks lead to biases in the cosmological parameters, greater than the credible interval. Thus the assembly bias is indeed the most dangerous systematic effect, which violates the scaling relation of halo bias with halo mass. However, we note that, in this paper, the result is considered as the worst-case scenario, because we include the maximum possible effect of the assembly bias in the `assembly-b-ext` mock, where we populate galaxies into halos assuming a fully deterministic assignment of centrals in ascending order of halo concentration: we populate galaxies from the lowest concentration halos in each mass bin. The impact of the assembly bias for more realistic galaxies, even if exists for SDSS galaxies, would be much smaller than what is shown in this paper. Nevertheless, this possible impact of the assembly bias needs to be kept in mind.
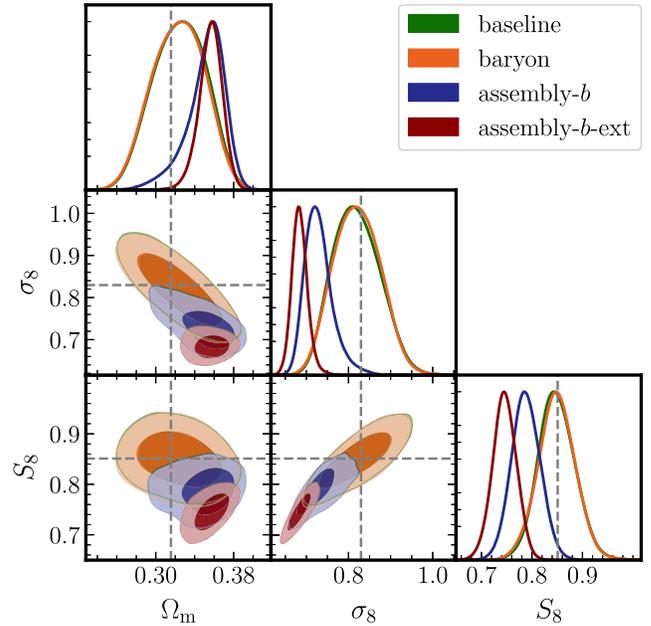


FIG. 17. Similar to the previous figure, but shown is the marginalized posterior distributions for the `baryon` mocks and the assembly bias mocks (`assembly-b` and `assembly-b-ext`). Here the `assembly-b-ext` is the worst (extreme) effect of assembly bias, where the large-scale amplitude of $w_{\mathrm{p}}$ is modified by a factor of 1.6 (see Fig. 4). In the `assembly-b` mock, the boost in the amplitude is halved, which is still larger than what is expected for an actual data, even if exists.
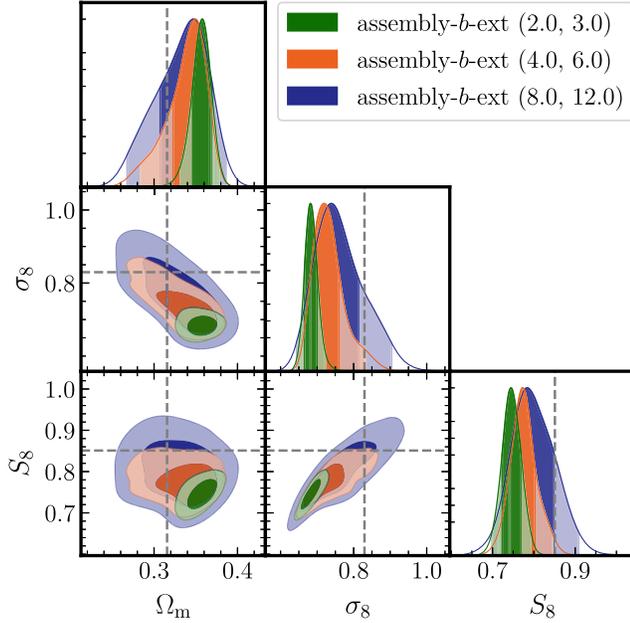
FIG. 18. Similar to the previous figure, but we here study how biases in the cosmological parameters can be mitigated by using the different scale cuts, (4,6) or (8, 12) $h^{-1}$ Mpc for $w_{\rm p}$ and $\Delta\Sigma$, respectively, instead of the baseline setup (2,3). The true cosmological parameters are recovered if using the large-scale cuts such as (8,12); that is, if we do not include the 1-halo term contribution.

To tackle the possibility that the actual galaxy sample is suffering from the assembly bias effect, we here propose a practical solution to assess its impact or mitigate it in parameter estimation. In Fig. 18, we show the results when using different scale cuts, (4,6) or (8, 12) $h^{-1}$ Mpc, for $w_{\rm p}$ and $\Delta\Sigma$, instead of our fiducial choice (2,3). Note that (8,12) is the choice of the minimal bias method in Sugiyama *et al.* [26] (also see Figs. 11 and 14). The figure clearly shows that the posterior distribution systematically moves with varying the scale cuts, and then the choice of (8,12) recovers the true cosmological parameters. In this case, the method does not use the 1-halo term signal of $\Delta\Sigma$, and tries to fit the mock signals by the model templates, where $r_{\rm cc} \simeq 1$ is satisfied. Thus, if we observe a similar systematic shift in the parameter constraints with varying the scale cuts, it could be a bias due to assembly bias. Although such a shift can be caused by other observational effects, observing this signature means that we need to investigate the origin of the shift before unblinding an analysis.

## VI. DISCUSSION AND CONCLUSION

In this paper we have in detail studied the performance of the halo model based method for cosmological parameter estimation. We used DARK EMULATOR to model the halo clustering quantities (halo mass function, halo

autocorrelation function, and halo-matter cross correlation), where the emulator includes, by design, all nontrivial effects such as nonlinear clustering, nonlinear halo bias and halo exclusion effect that are otherwise difficult to analytically model. We combined DARK EMULATOR with the HOD method to model clustering observables of galaxies for which we consider the projected correlation function, $w_{\rm p}(R)$, and the galaxy-galaxy weak lensing $\Delta\Sigma(R)$ in this paper. Then we validate the emulator-based halo model method by studying whether to recover the cosmological parameters from MCMC analyses by comparing the model predictions with the mock signals for the spectroscopic SDSS galaxies and the HSC-Y1 galaxies.

The main results of this paper are summarized as follows:

(1) Our method using DARK EMULATOR allows for computations of $w_{\rm p}$ and $\Delta\Sigma$ at a few CPU seconds for each model, which is equivalent to a factor of million reduction in computation time compared to the standard method (run $N$-body simulations, populate galaxies in halos, and then measure the galaxy clustering observables from the mocks). With this emulator, we can perform the MCMC analysis in practice.

(2) Changes in the cosmological parameters cause characteristic changes in the amplitudes and scale dependences of $w_{\rm p}$ and $\Delta\Sigma$ via a combination of various effects: nonlinear clustering, nonlinear halo bias and halo exclusion effect (Fig. 2).

(3) We showed that the halo model based method can recover the underlying cosmological parameters, especially $S_8 = \sigma_8\Omega_{\rm m0}^{0.5}$, after marginalization over the galaxy-halo connection (HOD) parameters, for a variety of mock catalogs except for the mocks including an extreme effect of assembly bias, for the nominal choices of scale cuts, $(R_{\rm cut}^{w_{\rm p}}, R_{\rm cut}^{\Delta\Sigma}) = (2, 3)h^{-1}$ Mpc (Figs. 12 and 14). The baseline method can achieve a precision of $\sigma(S_8) \simeq$ 0.035–0.042 for the SDSS and HSC-Y1 datasets. This method allows for tight constraints on the cosmological parameter, because the small-scale information of $\Delta\Sigma(R)$ at scales, $3 \leq R/[h^{-1}\,{\rm Mpc}] \lesssim 10$, can be used to infer the average mass of host halos of the SDSS galaxies, yielding useful information on the galaxy bias that is sensitive to the large-scale amplitudes of $w_{\rm p}$ and $\Delta\Sigma$.

(4) The halo model method suffers from sizable biases in the cosmological parameters if the SDSS galaxies have a significant assembly bias (although we consider the extreme mocks including the maximum amount of the assembly bias effect in this paper) (Figs. 14 and 17). Even if this is the case, we showed that the cosmological parameters can be recovered if employing sufficiently large scales cuts of $R_{\rm cut} \gtrsim$ $10\ h^{-1}$ Mpc, where the cross-correlation coefficient

$r_{cc}(r) \equiv \xi_{gm}(r)/[\xi_{gg}(r)\xi_{mm}(r)]^{1/2} \simeq 1$  (Fig. 18). This is equivalent to the case that we do not include the small-scale lensing information in the cosmological parameter estimation. However the price to pay is that the statistical accuracy of cosmological parameter determination is degraded, leading to $\sigma(S_8) \simeq 0.07$ (see Fig. 14). Thus in practice we should monitor whether the cosmological parameters have a systematic shift with different scale cuts, as a test of the assembly bias effect inherent in actual data (Fig. 17).

(5) We used the method using a single population of source galaxies to measure the galaxy-galaxy weak lensing for multiple lens galaxies at different redshifts (LOWZ, CMASS1, and CMASS2) over the range of redshifts, $z = [0.15, 0.7]$. This method allows for self-calibration of the photometric redshift errors and the multiplicative shear bias (Fig. 12).

Hence we conclude that we can safely apply the halo model based method to actual SDSS and HSC-Y1 data. The cosmological constraints from actual observational data are presented in a following paper [45].

## ACKNOWLEDGMENTS

## APPENDIX A: COSMOLOGICAL DEPENDENCE OF THE 2-HALO TERM OF GALAXY CORRELATION FUNCTIONS

The cosmological information in the joint probes cosmology using $\Delta\Sigma$ and $w_p$ lies in the 2-halo terms of $\Delta\Sigma$ and $w_p$. In particular, an advantage in the use of DARK EMULATOR is that it includes all the nonlinear effects such as the nonlinear matter clustering, the nonlinear halo bias and the halo exclusion effect that are very difficult to accurately calibrate unless $N$-body simulations are employed as done in DARK EMULATOR. In this appendix, we study how these nonlinear effects cause scale-dependent variations in the 2-halo term by changes in the cosmological parameters.

Figure 19 shows how a change in $\sigma_8$ or $\Omega_{m0}$ causes a scale-dependent change in the 2-halo term of the three-dimensional correlation function, $\xi_{gg,2h}$, while other parameters including the HOD parameters are kept to their fiducial values. Note that we use DARK EMULATOR to compute these results, and we here focus on the galaxy autocorrelation function because most of the cosmological
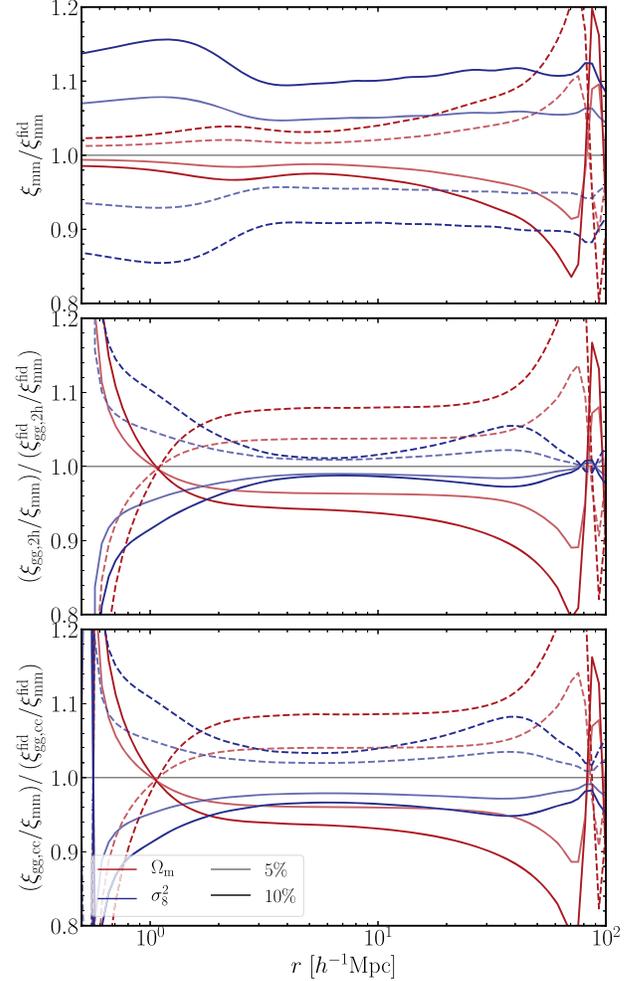


FIG. 19. Fractional changes in the 2-halo term of the three-dimensional galaxy-galaxy correlation function relative to that for the fiducial model, when varying either of $\sigma_8$ or $\Omega_{m0}$. Note that other model parameters keep fixed to their fiducial values. The middle panel shows the fractional change for the ratio $\xi_{gg,2h}/\xi_{mm}$, because the ratio becomes scale independent if the linear theory holds, which predicts $\xi_{gg} = b^2 \xi_{mm}$ with a scale-independent coefficient $b$. The lower panel shows the result for the two-point correlation function of central galaxies that includes only the 2-halo term by definition. The different lines show the results when $\pm 5\%$ and $\pm 10\%$ changes in $\sigma_8^2$ and $\Omega_{m0}$, and the solid or dashed lines correspond to the results when increasing or decreasing the parameter by the amount from the fiducial value. Here we consider the HOD for LOWZ-like galaxies at $z = 0.25$.

information for the SDSS and HSC datasets are from $w_p$, which is obtained from the projection of $\xi_{gg}$. First, for comparison, the upper panel shows the change in the matter correlation function, $\xi_{mm}$, relative to that for the fiducial model (*Planck* cosmology). The figure shows that a change in $\sigma_8$ causes a fairly sale-independent change in $\xi_{mm}$ at scales $r \lesssim 70 \ h^{-1}$ Mpc, which we consider throughout this paper. On the other hand, a change in $\Omega_{m0}$ causes a scale-dependent change in $\xi_{mm}$.

The middle panel shows the fractional change in $\xi_{gg,2h}/\xi_{mm}$ relative to the ratio for the fiducial model. Note that we here consider the fiducial model of HOD for LOWZ-like galaxies at $z = 0.25$. The reason we consider this ratio is that, if $\xi_{gg}$ follows the linear theory prediction as given by $\xi_{gg} = b^2 \xi_{mm}$ with a scale-independent coefficient $b$, then the change becomes scale independent for the change in the cosmological parameters. Here, for the sake of clarity, we consider only the 2-halo term of $\xi_{gg}$ using Eq. (16). The figure shows that a change in $\sigma_8$ causes a scale-dependent change in the ratio, which should arise from the scale-dependent change in the halo bias and the halo exclusion effect by the change in $\sigma_8$. The change in $\Omega_{m0}$ is also found to cause a scale-dependent change in the ratio, again via the dependence of halo bias on $\Omega_{m0}$. Thus DARK EMULATOR includes these complex dependences of $\xi_{gg}$ on the cosmological parameters. For further comparison, the lower panel shows the results for $\xi_{gg,cc}$, i.e., the two-point correlation function of central galaxies, which includes only the 2-halo term by definition. The results appear to be similar to those in the middle panel.

Thus DARK EMULATOR includes these complex dependences of the galaxy-galaxy correlation function on the cosmological parameters ($\sigma_8$ and $\Omega_{m0}$). These dependences are difficult to accurately calibrate analytically, e.g., by the perturbation theory, and the accurate calibration requires the use of $N$-body simulations as done in the development of DARK EMULATOR. These complex cosmological dependences of $\xi_{gg}$ are an advantage of our method.

## APPENDIX B: COVARIANCE ESTIMATION BASED ON MOCK CATALOGS

In this appendix, we describe the estimation of statistical uncertainties in the galaxy-galaxy weak lensing profile ($\Delta\Sigma$) by using a set of synthetic observational datasets, i.e., the covariance matrix.

A robust estimation of the covariance matrix is one of the most important subjects in modern cosmology in practice (e.g., see Refs. [101–105]). In this paper, we use a set of numerical simulations in Ref. [79] to construct realistic mock catalogs of galaxy shapes as well as the tracers of large-scale structures. We then adopt the same analysis pipeline to measure $\Delta\Sigma$ from the mock catalogs as we do for actual measurements. Our mock measurements of the lensing signals have 2268 realizations in total, allowing us to evaluate the covariance matrix of $\Delta\Sigma$ for the SDSS-like

galaxies at multiple redshifts in a rigorous way. In the following, we describe how to produce mock catalogs from the simulations incorporated with observational data and show the validation of our mock lensing analyses.

### 1. Massive production of mock catalogs

#### a. Full-sky simulations of lensing and halos

We first briefly introduce the full-sky ray-tracing simulations and the halo catalogs in the line-cone simulation realization, developed in Ref. [79] (the full-sky simulation data are freely available from [106]). The full-sky simulations are based on a set of $N$-body simulations with $2048^3$ particles in cosmological volumes. Reference [79] adopted the standard $\Lambda$CDM cosmology with the following cosmological parameters: the CDM density parameter $\Omega_{cdm} = 0.233$, the baryon density $\Omega_b = 0.046$, the matter density $\Omega_m = \Omega_{cdm} + \Omega_b = 0.279$, the cosmological constant $\Omega_\Lambda = 0.721$, the Hubble parameter $h = 0.7$, the amplitude of density fluctuations $\sigma_8 = 0.82$, and the spectral index $n_s = 0.97$. Note that the cosmological model in the simulation is consistent with the Wilkinson microwave anisotropy probe nine-year cosmology [107].

Full-sky weak gravitational lensing simulations have been performed with the standard multiple lens-plane algorithm [108–110]. In this simulation, one can take into account the light-ray deflection on the celestial sphere by using the projected matter density field given in the format of spherical shell (see the similar approach for Ref. [111]). The simulations used the projected matter fields in 38 shells in total, each of which was computed by projecting $N$-body simulation realization over a radial width of $150 \ h^{-1}$ Mpc, in order to make the light cone simulation covering a cosmological volume up to $z = 5.3$. As a result, the lensing simulations consist of the shear field at each of 38 different source redshifts with angular resolution of 0.43 arcmin. Each simulation data is given in the HEALPIX format [112]. The radial depth between nearest source redshifts is set to be $150 \ h^{-1}$ Mpc in comoving distance, corresponding to the redshift interval of 0.05–0.1 for $z \lesssim 1$.

In each output of the $N$-body simulation, Ref. [79] identified dark matter halos using the ROCKSTAR algorithm [73]. In the following, we define the halo mass by using the spherical overdensity criterion: $M_{200\,m} = 200\bar{\rho}_{m0}(4\pi/3)R_{200\,m}^3$. Individual halos in $N$-body boxes are assigned to the pixels in the celestial sphere with the HEALPIX software. It should be noted that the $N$-body simulations allow us to resolve dark matter halos with masses greater than a few times $10^{12}h^{-1} \ M_\odot$ with more than $50N$-body particles at redshifts $z < 0.7$, which is a typical redshift range of massive galaxies in the SDSS BOSS survey.

#### b. Shapes of background galaxies

For shapes of background galaxies, we use the mock catalogs produced in Ref. [81]. The mocks are specific to

cosmological analyses with the Subaru HSC-Y1 shape catalog [113] by taking into account various observational effects as the survey footprints, inhomogeneous angular distribution of source galaxies, statistical uncertainties in photometric redshift estimate, variations in the lensing weight, and the statistical noise in galaxy shape measurements including both intrinsic shapes and the measurement errors. We produced 2268 mock catalogs from 108 full-sky ray-tracing simulations of gravitational lensing in Ref. [79] by using a similar approach developed in Refs. [80,114]. We properly incorporated with the simulated lensing shear and the observed galaxy shape on an object-by-object basis, enabling the mock to share exactly the same information of angular positions, redshifts, the lensing weights and the shear responsivity with the real catalog. The further details are found in Ref. [81]. The mock shape catalogs are publicly available at [115].

### c. Mock catalogs of lensing galaxies, SDSS-like galaxies

For the foreground galaxies to be used in the galaxy-galaxy lensing analysis, we produce the mock galaxy catalogs assuming the HOD method as summarized in Sec. III B. As in the mock galaxy shape catalogs, we first extract 2268 realizations of the HSC-Y1 survey windows from 108 full-sky halo catalogs in Ref. [79]. We then populate galaxies into halos using the fiducial HOD method whose parameters are chosen to mimic LOWZ and CMASS galaxies in the redshift range of $0.15 < z < 0.35$, $0.43 < z < 0.55$, and $0.55 < z < 0.70$, but spanning the entire SDSS BOSS footprint. The HOD method is used to populate mock LOWZ and CMASS galaxies in halos of each of the light cone simulation realization.

Besides, we include the redshift-space distortion effects in mock LOWZ and CMASS galaxies. Along the line of sight of each host halo, we set the radial velocity of the central galaxy to be the same as one of its host halo, while we assign the random velocities to the satellites by following a Gaussian distribution with width given by the virial velocity dispersion.

### 2. Validation of our covariance estimation

In this section, we validate the performance of our covariance estimation with the mock measurements of galaxy-galaxy lensing signals. For comparison, we predict the Gaussian covariances of the lensing signals by a halo-model framework developed in Ref. [69]. In this framework, we can properly take into account the weight function related to the conversion between the lensing shear and the excess surface mass density in the covariance matrix. It would be worth noting that the covariance model has been extensively validated with a set of numerical simulations for a variety of the weight functions in the lensing analysis (see Ref. [69] for details).

Figure 20 shows the comparison of the variance of the lensing signal in each of three redshift bins with the simulation results and their Gaussian predictions. In each panel, the simulation results are shown in the colored points, while the solid and dashed lines represent the Gaussian covariances with and without the shape noise, respectively. The figure clearly shows that the sample variance caused by the line-of-sight large scale structures dominates the mock variance, and the shot noise is dominant at the length scale less than $10\ h^{-1}$ Mpc.

We then study the off-diagonal elements of the mock covariance matrices. The left panels in Fig. 21 summarize the comparison of the cross correlation coefficient in the covariance for the lensing signals at the lens redshift $0.43 < z < 0.55$. This figure shows that our simple Gaussian prediction provides a reasonable agreement with the off-diagonal elements of the mock covariance,
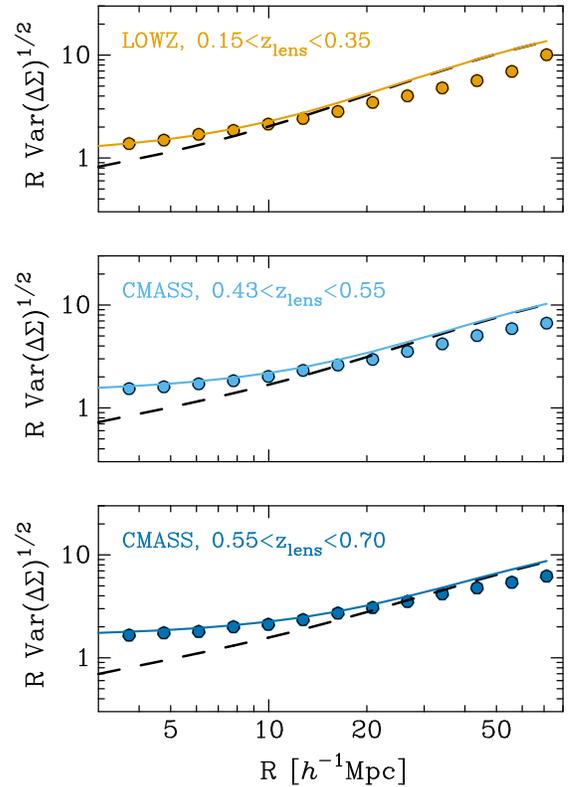


FIG. 20. Comparison of the diagonal components in the covariance matrices for the galaxy-galaxy lensing signals. From top to bottom, we show the variances of the lensing signals at three different redshift bins (the lower panel corresponds to the case at the higher redshift). The points represent the simulation results based on 2268 realizations, while the solid lines stand for the Gaussian predictions developed in Ref. [69]. In each panel, the dashed line shows the variance in the absence of shape noises, highlighting the contribution from uncorrelated large-scale structures along a line of the sight. Note that we define $\Delta\Sigma$ in units of $h\,M_\odot\mathrm{pc}^{-2}$ in this figure. Hence, the unit in each vertical axis is given by $\mathrm{Mpc}(M_\odot\,\mathrm{pc}^{-2})$.
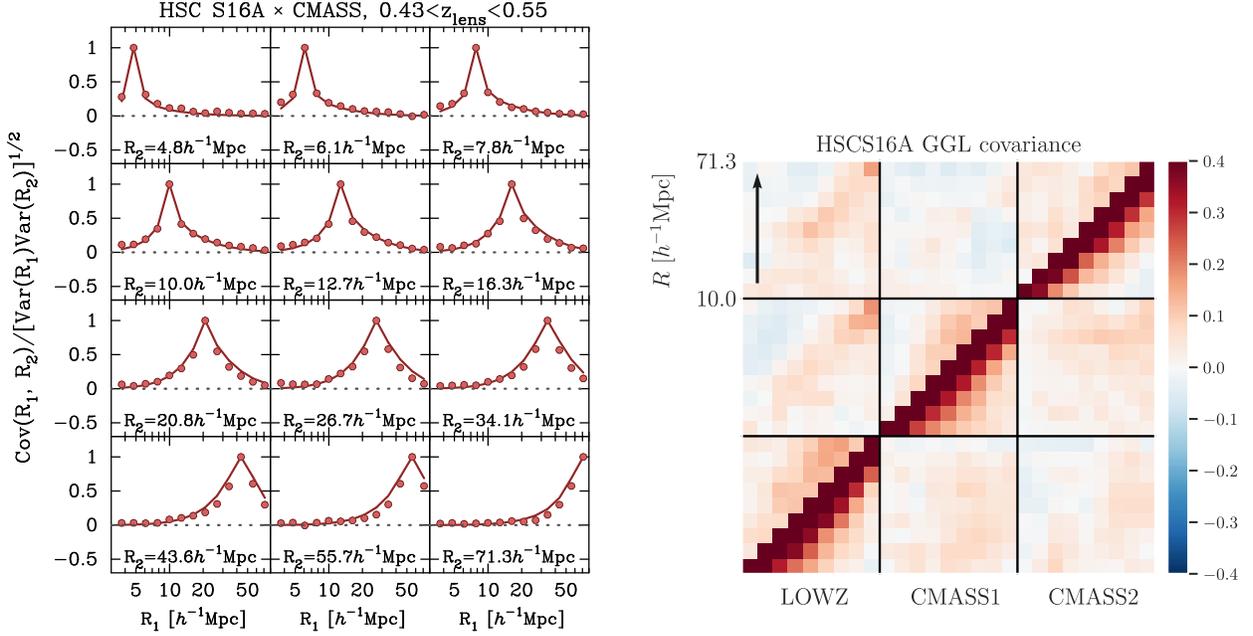
FIG. 21. Comparisons of the off-diagonal elements in the covariance matrices with the simulation results and their Gaussian predictions. The left panels show the cross correlation coefficient in the covariance at the medium redshift bin (in the redshift range of $0.43 < z < 0.55$) as a function of radii. In the left, the points show the simulation results, while the lines are for the Gaussian predictions. Each small panel in the left represents the scale dependence of the cross correlation coefficient. In the right, we show the cross correlation coefficients in the full covariance across three redshift bins. The lower triangular panels show the simulation results, while the upper one displays the difference between the simulation results and the Gaussian predictions. In the right, we use the labels of "LOWZ," "CMASS1," and "CMASS2" from the lowest to the highest redshift bins. Note that we limit the length scales to be in the range of $R > 10 \ h^{-1}$ Mpc in the right panels to focus on the sample covariance. The diagonal elements in the right is set to 1 by construction.

indicating the super-sample covariance (SSC) [116] is not important to our galaxy-galaxy lensing analyses. The SSC is expected to arise from the four-point correlation among super- and subsurvey modes [116,117], and the recent simulations have shown that the SSC in the halo-matter cross correlation becomes important only at $1h^{-1}$ Mpc [118]. At the scale of $R \sim 1 \ h^{-1}$ Mpc, the statistical uncertainties in our lensing analyses are mostly determined by the shot noise terms.

The right panel of Fig. 21 shows the cross correlation coefficient of the mock covariance matrix across three redshift bins. The lower triangular panel shows the simulation results, while the upper panels represent the difference between the simulation results and the Gaussian predictions. We find the large cross correlation coefficients in the radius range of $R > 10h^{-1}$ Mpc with a level of ∼0.5 at most for single redshifts and the cross covariance

between two different redshifts is less prominent. Our Gaussian covariance is in good agreement with the simulation results, allowing us to explain the mock covariance with a 20% level accuracy.

## APPENDIX C: POSTERIOR DISTRIBUTION OF PARAMETERS IN A FULL-DIMENSION PARAMETER SPACE

For comprehensiveness of our discussion, in Fig. 22 we show the posterior distribution for all the parameters including all the HOD parameters for the CMASS1-like galaxies. The results for the LOWZ- and CMASS2-like galaxies at different redshifts are similar to this plot. Even if we use the joint information of $w_p$ and $\Delta\Sigma$, each of the HOD parameters is not well constrained, but the cosmological parameters are recovered.
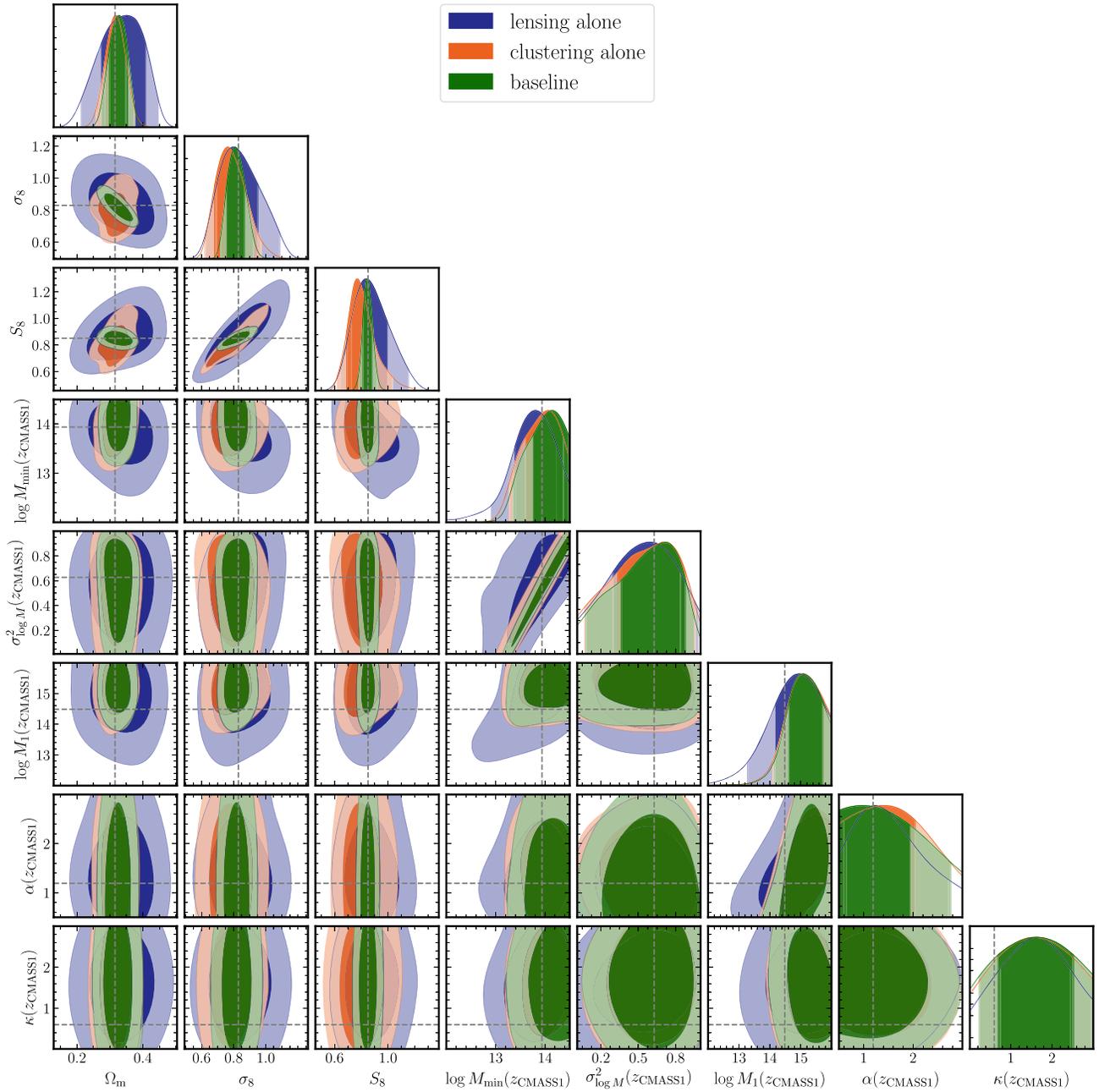
FIG. 22. Posterior distributions of cosmological parameters and HOD parameters, as in Fig. 9. Here we show the results for all the HOD parameters for the CMASS1-like galaxies (see Table II).

[1] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo, Phys. Rep. **530,** 87 (2013).

[2] https://hsc.mtk.nao.ac.jp/ssp/.

[3] H. Aihara *et al.*, Publ. Astron. Soc. Jpn. **70,** S4 (2018).

[4] C. Hikage *et al.*, Publ. Astron. Soc. Jpn. **71,** 43 (2019).

[5] T. Hamana *et al.*, Publ. Astron. Soc. Jpn. **72,** 16 (2020).

[6] https://www.darkenergysurvey.org.

[7] T. M. C. Abbott *et al.* (Dark Energy Survey Collaboration), Phys. Rev. D **98,** 043526 (2018).

[8] C. Heymans *et al.*, Astron. Astrophys. **646,** A140 (2021).

[9] Y. Park and E. Rozo, Mon. Not. R. Astron. Soc. **499,** 4638 (2020).

[10] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **641,** A6 (2020).

[11] https://pfs.ipmu.jp/index.html.

[12] M. Takada *et al.*, Publ. Astron. Soc. Jpn. **66,** R1 (2014).

[13] https://www.desi.lbl.gov.

[14] https://www.lsst.org.

[15] https://www.euclid-ec.org.

[16] https://roman.gsfc.nasa.gov.

[17] D. Spergel *et al.*, arXiv:1503.03757.

[18] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist, Nature (London) **509,** 177 (2014).

[19] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman, Mon. Not. R. Astron. Soc. **475,** 676 (2018).

[20] U. Seljak *et al.*, Phys. Rev. D **71,** 103515 (2005).

[21] M. Oguri and M. Takada, Phys. Rev. D **83,** 023008 (2011).

[22] M. Cacciato, F. C. van den Bosch, S. More, H. Mo, and X. Yang, Mon. Not. R. Astron. Soc. **430,** 767 (2013).

[23] R. Mandelbaum, A. Slosar, T. Baldauf, U. Seljak, C. M. Hirata, R. Nakajima, R. Reyes, and R. E. Smith, Mon. Not. R. Astron. Soc. **432,** 1544 (2013).

[24] H. Miyatake, S. More, R. Mandelbaum, M. Takada, D. N. Spergel, J.-P. Kneib, D. P. Schneider, J. Brinkmann, and J. R. Brownstein, Astrophys. J. **806,** 1 (2015).

[25] S. More, H. Miyatake, R. Mandelbaum, M. Takada, D. N. Spergel, J. R. Brownstein, and D. P. Schneider, Astrophys. J. **806,** 2 (2015).

[26] S. Sugiyama, M. Takada, Y. Kobayashi, H. Miyatake, M. Shirasaki, T. Nishimichi, and Y. Park, Phys. Rev. D **102,** 083520 (2020).

[27] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367,** 1 (2002).

[28] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rep. **733,** 1 (2018).

[29] E. Krause *et al.*, arXiv:1706.09359.

[30] N. MacCrann *et al.* (DES Collaboration), Mon. Not. R. Astron. Soc. **480,** 4614 (2018).

[31] T. Nishimichi, G. D'Amico, M. M. Ivanov, L. Senatore, M. Simonović, M. Takada, M. Zaldarriaga, and P. Zhang, Phys. Rev. D **102,** 123541 (2020).

[32] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, 1980), https://ui.adsabs .harvard.edu/abs/1980lssu.book.....P/abstract.

[33] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 07 (2012) 051.

[34] U. Seljak, Mon. Not. R. Astron. Soc. **318,** 203 (2000).

[35] J. A. Peacock and R. E. Smith, Mon. Not. R. Astron. Soc. **318,** 1144 (2000).

[36] C. Ma and J. N. Fry, Astrophys. J. **543,** 503 (2000).

[37] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain, Astrophys. J. **546,** 20 (2001).

[38] A. Cooray and R. Sheth, Phys. Rep. **372,** 1 (2002).

[39] Y. P. Jing, H. J. Mo, and G. Börner, Astrophys. J. **494,** 1 (1998).

[40] Z. Zheng *et al.*, Astrophys. J. **633,** 791 (2005).

[41] T. Nishimichi, M. Takada, R. Takahashi, K. Osato, M. Shirasaki, T. Oogi, H. Miyatake, M. Oguri, R. Murata, Y. Kobayashi, and N. Yoshida, Astrophys. J. **884,** 29 (2019).

[42] S. Alam *et al.*, Mon. Not. R. Astron. Soc. **470,** 2617 (2017).

[43] R. Mandelbaum *et al.*, Publ. Astron. Soc. Jpn. **70,** S25 (2018).

[44] Y. Kobayashi, T. Nishimichi, M. Takada, and R. Takahashi, Phys. Rev. D **101,** 023510 (2020).

[45] H. Miyatake *et al.*, following paper, Phys. Rev. D **106,** 083520 (2022).

[46] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594,** A13 (2016).

[47] R. Mandelbaum, U. Seljak, G. Kauffmann, C. M. Hirata, and J. Brinkmann, Mon. Not. R. Astron. Soc. **368,** 715 (2006).

[48] A. Schneider and R. Teyssier, J. Cosmol. Astropart. Phys. 12 (2015) 049.

[49] A. Schneider, R. Teyssier, J. Stadel, N. E. Chisari, A. M. C. Le Brun, A. Amara, and A. Refregier, J. Cosmol. Astropart. Phys. 03 (2019) 020.

[50] F. C. van den Bosch, S. More, M. Cacciato, H. Mo, and X. Yang, Mon. Not. R. Astron. Soc. **430,** 725 (2013).

[51] Y. Kobayashi, T. Nishimichi, M. Takada, R. Takahashi, and K. Osato, Phys. Rev. D **102,** 063504 (2020).

[52] Y. Kobayashi, T. Nishimichi, M. Takada, and H. Miyatake, Phys. Rev. D **105,** 083517 (2022).

[53] https://jila.colorado.edu/~ajsh/FFTLog/index.html.

[54] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490,** 493 (1997).

[55] B. Diemer and A. V. Kravtsov, Astrophys. J. **799,** 108 (2015).

[56] https://bdiemer. bitbucket.io/colossus/.

[57] B. Diemer, Astrophys. J. Suppl. **239,** 35 (2018).

[58] C. Hikage, R. Mandelbaum, M. Takada, and D. N. Spergel, Mon. Not. R. Astron. Soc. **435,** 2345 (2013).

[59] S. Masaki, C. Hikage, M. Takada, D. N. Spergel, and N. Sugiyama, Mon. Not. R. Astron. Soc. **433,** 3506 (2013).

[60] I. Mohammed and U. Seljak, Mon. Not. R. Astron. Soc. **445,** 3382 (2014).

[61] O. H. E. Philcox, D. N. Spergel, and F. Villaescusa-Navarro, Phys. Rev. D **101,** 123520 (2020).

[62] B. Hadzhiyska, S. Bose, D. Eisenstein, and L. Hernquist, Mon. Not. R. Astron. Soc. **501,** 1603 (2020).

[63] Due to the line-of-sight projection, the BAO features, which are around 100 $h^{-1}$ Mpc in the three-dimensional correlation function, appear at a shorter projected separation.

[64] S. More, Astrophys. J. Lett. **777,** L26 (2013).

[65] M. Tanaka, J. Coupon, B.-C. Hsieh, S. Mineo, A. J. Nishizawa, J. Speagle, H. Furusawa, S. Miyazaki, and H. Murayama, Publ. Astron. Soc. Jpn. **70,** S9 (2018).

[66] G. M. Bernstein and M. Jarvis, Astron. J. **123,** 583 (2002).

[67] E. S. Sheldon, D. E. Johnston, J. A. Frieman, R. Scranton, T. A. McKay, A. J. Connolly, T. Budavári, I. Zehavi, N. A. Bahcall, J. Brinkmann, and M. Fukugita, Astron. J. **127,** 2544 (2004).

[68] R. Mandelbaum, A. Tasitsiomi, U. Seljak, A. V. Kravtsov, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **362,** 1451 (2005).

[69] M. Shirasaki and M. Takada, Mon. Not. R. Astron. Soc. **478,** 4277 (2018).

[70] D. Huterer, M. Takada, G. Bernstein, and B. Jain, Mon. Not. R. Astron. Soc. **366,** 101 (2006).

[71] S. Bridle *et al.*, Mon. Not. R. Astron. Soc. **405,** 2044 (2010).

[72] N. Kaiser, Mon. Not. R. Astron. Soc. **227,** 1 (1987).

[73] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, Astrophys. J. **762,** 109 (2013).

[74] https://www.sdss.org/dr11/.

[75] A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson, Astrophys. J. **499,** 20 (1998).

[76] P. Valageas and T. Nishimichi, Astron. Astrophys. **527,** A87 (2011).

[77] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Astrophys. J. **761,** 152 (2012).

[78] We use the slightly different number of realizations for $\Delta\Sigma$ and $w_p$, because the *N*-body simulation data for the four realizations in the difference were lost in the middle of this work. The difference is small compared to the statistical measurement errors for the SDSS and HSC data, and the main results of this paper are not changed.

[79] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama, Astrophys. J. **850,** 24 (2017).

[80] M. Shirasaki, M. Takada, H. Miyatake, R. Takahashi, T. Hamana, T. Nishimichi, and R. Murata, Mon. Not. R. Astron. Soc. **470,** 3476 (2017).

[81] M. Shirasaki, T. Hamana, M. Takada, R. Takahashi, and H. Miyatake, Mon. Not. R. Astron. Soc. **486,** 52 (2019).

[82] D. Kodwani, D. Alonso, and P. Ferreira, Open J. Astrophys. **2,** 3 (2019).

[83] Precisely speaking, the binning of *R* does not have an exact cut at 2, 3, or 30 $h^{-1}$ Mpc for the lower and upper cuts. We employ the 30 bins logarithmically spaced over $0.05 \leq R/[h^{-1} \text{ Mpc}] \leq 80$, and use the bins residing the fitting range of $\Delta\Sigma$ and $w_p$.

[84] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398,** 1601 (2009).

[85] B. Audren, J. Lesgourgues, K. Benabed, and S. Prunet, J. Cosmol. Astropart. Phys. 02 (2013) 001.

[86] B. Joachimi *et al.*, Astron. Astrophys. **646,** A129 (2021).

[87] L. Gao, V. Springel, and S. D. M. White, Mon. Not. R. Astron. Soc. **363,** L66 (2005).

[88] R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood, Astrophys. J. **652,** 71 (2006).

[89] N. Dalal, M. White, J. R. Bond, and A. Shirokov, Astrophys. J. **687,** 12 (2008).

[90] E. Semboloni, H. Hoekstra, and J. Schaye, Mon. Not. R. Astron. Soc. **434,** 148 (2013).

[91] A. Leauthaud *et al.*, Mon. Not. R. Astron. Soc. **467,** 3024 (2017).

[92] The code of analytical halo model is available from https://github.com/surhudm/aum.

[93] J. L. Tinker, D. H. Weinberg, Z. Zheng, and I. Zehavi, Astrophys. J. **631,** 41 (2005).

[94] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber, Astrophys. J. **724,** 878 (2010).

[95] Y. Li, W. Hu, and M. Takada, Phys. Rev. D **93,** 063507 (2016).

[96] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 09 (2016) 007.

[97] T. Lazeyras, C. Wagner, T. Baldauf, and F. Schmidt, J. Cosmol. Astropart. Phys. 02 (2016) 018.

[98] M. Oguri *et al.*, Publ. Astron. Soc. Jpn. **70,** S20 (2018).

[99] H. Miyatake *et al.*, Astrophys. J. **875,** 63 (2019).

[100] R. Mandelbaum, F. Lanusse, A. Leauthaud, R. Armstrong, M. Simet, H. Miyatake, J. E. Meyers, J. Bosch, R. Murata, S. Miyazaki, and M. Tanaka, Mon. Not. R. Astron. Soc. **481,** 3170 (2018).

[101] J. Hartlap, P. Simon, and P. Schneider, Astron. Astrophys. **464,** 399 (2007).

[102] A. Taylor, B. Joachimi, and T. Kitching, Mon. Not. R. Astron. Soc. **432,** 1928 (2013).

[103] S. Dodelson and M. D. Schneider, Phys. Rev. D **88,** 063537 (2013).

[104] D. J. Paz and A. G. Sánchez, Mon. Not. R. Astron. Soc. **454,** 4326 (2015).

[105] O. Friedrich and T. Eifler, Mon. Not. R. Astron. Soc. **473,** 4150 (2018).

[106] http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/.

[107] G. Hinshaw *et al.*, Astrophys. J. Suppl. Ser. **208,** 19 (2013).

[108] T. Hamana and Y. Mellier, Mon. Not. R. Astron. Soc. **327,** 169 (2001).

[109] M. R. Becker, Mon. Not. R. Astron. Soc. **435,** 115 (2013).

[110] M. Shirasaki, T. Hamana, and N. Yoshida, Mon. Not. R. Astron. Soc. **453,** 3044 (2015).

[111] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Manera, Mon. Not. R. Astron. Soc. **391,** 435 (2008).

[112] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622,** 759 (2005).

[113] R. Mandelbaum *et al.*, Publ. Astron. Soc. Jpn. **70,** S25 (2018).

[114] M. Shirasaki and N. Yoshida, Astrophys. J. **786,** 43 (2014).

[115] http://gfarm.ipmu.jp/~surhud/.

[116] M. Takada and W. Hu, Phys. Rev. D **87,** 123504 (2013).

[117] Y. Li, W. Hu, and M. Takada, Phys. Rev. D **89,** 083519 (2014).

[118] R. Takahashi, T. Nishimichi, M. Takada, M. Shirasaki, and K. Shiroyama, Mon. Not. R. Astron. Soc. **482,** 4253 (2019).