

# Galaxy Clustering Constraints on the Galaxy Size–Halo Connection

Andrew Hearin<sup>1</sup>, Peter Behroozi<sup>2</sup>, Andrey Kravtsov<sup>3</sup>, Benjamin Moster<sup>4</sup>

<sup>1</sup>*Argonne National Laboratory, Argonne, IL, USA 60439, USA*

<sup>2</sup>*Department of Physics, University of Arizona, 1118 E 4th St, Tucson, AZ 85721 USA*

<sup>3</sup>*Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637 USA*

<sup>4</sup>*Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany*

2 September 2017

## ABSTRACT

We derive empirical modeling constraints on the connection between dark matter halos and the half-mass radius  $R_{1/2}$  of galaxy bulges and disks. We show that both  $R_{1/2}^{\text{disk}}$  and  $R_{1/2}^{\text{bulge}}$  are well-described by power law scaling relations with halo virial radius,  $R_{1/2} = AR_{\text{vir}}^\alpha$ . Novel to this work, we use new SDSS measurements of the  $R_{1/2}$ –dependence of galaxy clustering to constrain the model parameters,  $A_{\text{bulge}}$ ,  $A_{\text{disk}}$ ,  $\alpha_{\text{bulge}}$ ,  $\alpha_{\text{disk}}$ , and log-normal scatter  $\sigma_{R_{1/2}}$ . Even when only coarsely tuning these parameters to the observed one-point functions  $\langle R_{1/2}^{\text{disk}} | M_*^{\text{disk}} \rangle$  and  $\langle R_{1/2}^{\text{bulge}} | M_*^{\text{bulge}} \rangle$ , our model accurately predicts the observed two-point clustering on small- and large-scales. This success non-trivial, as we show that galaxy clustering is highly sensitive to the physics that shapes satellite galaxy profiles. We find no evidence for the commonly assumed relation between halo spin  $\lambda_{\text{halo}}$  and  $R_{1/2}^{\text{disk}}$ , and show that this assumption cannot be meaningfully constrained with either the clustering or lensing of  $L_*$  galaxies. Our results provide simple boundary conditions for more complex and fine-grained models of galaxy size. We make our python code publicly available to support cosmological surveys that require realistic synthetic galaxy populations.

## 1 INTRODUCTION

Some introduction goes here.

## 2 DATA AND SIMULATIONS

### 2.1 Galaxy Sample and Measurements

Our galaxy sample comes from Data Release 10 of the Sloan Digital Sky Survey (SDSS Ahn et al. 2014). We study the same  $M_*$ –complete galaxy sample used in Behroozi et al. (2015), to which we refer the reader for details. Values for  $M_*$  and SFR are taken from the MPA-JHU value-added catalog (Kauffmann et al. 2003; Brinchmann et al. 2004).

We additionally use the catalog of SDSS galaxy profile decompositions provided by Meert et al. (2015). In the version of this catalog that we use, two-dimensional  $r$ –band profiles were fit with a two-component de Vaucouleurs + exponential profile. We cross-match our DR10 sample against the Meert et al. (2015) catalog values of  $R_{1/2}$ , the half-light radius of the full profile, and B/T, the  $r$ –band flux ratio of the de Vaucouleurs component to the total  $r$ –band flux. This cross-matching gives us a volume-limited galaxy sample with values of  $M_*$ , sSFR,

$R_{1/2}$ , and B/T. Together with the present work, we make publicly available this cross-matched catalog, as well as code to generate the catalog from the original data files and load it into a python interpreter.

We calculate two-point clustering  $w_p$  of our SDSS galaxy sample using line-of-sight projection of  $\pi_{\text{max}} = 20\text{Mpc}$  using the `correl` program in `UniverseMachine`. Our results in § 4 will give special focus on the dependence of  $w_p$  upon composite galaxy size  $R_{1/2}$ . We will quantify this dependence in terms of *clustering ratios* of “large” vs. “small” galaxies, defined according to whether composite galaxy size is above or below  $\langle R_{1/2} | M_* \rangle$ , computed as the median of a sliding stellar mass window with a width of  $N_{\text{gal}} = 1000$ .

### 2.2 Simulation and Halos

`Rockstar` subhalos identified at  $z = 0$  in the Bolshoi-Planck simulation are the basis of all our modeling.

### 2.3 Baseline Model and Mock Measurements

#### 2.3.1 Mapping $M_*$ & SFR to halos

The starting point of our model is the best-fit **UniverseMachine** model (Behroozi et al. 2017, in prep). This model maps stellar mass and star-formation rate to every halo and subhalo at each snapshot in the Bolshoi-Planck simulation. The stellar mass function, quenched fraction, SFR density, and SFR-dependent clustering are all accurately captured by this model from  $z = 0-10$ . The model we have developed in the present work does not depend in an essential way upon the **UniverseMachine** model in particular, we only require some reasonably accurate starting point for the  $M_*$  and sSFR values mapped to simulated halos.

#### 2.3.2 Mapping B/T to halos

Our model for galaxy size is further predicated upon the mapping between dark matter halos and B/T, the fraction of stellar mass in the bulge. In a companion paper to this work, we will present a separate empirical model for the B/T–halo connection. For our present purposes, we instead rely upon the simplest possible assumption for this connection that recovers the observed distribution  $P(\text{B/T}|M_*, \text{sSFR})$ : we suppose that galaxy B/T has no dependence whatsoever upon halo environment or assembly, beyond what is inherited by the mutual correlation between  $M_*$ , sSFR and the cosmic web. That is, we map B/T to dark matter halos by using the  $M_*$  and sSFR values from **UniverseMachine** to randomly select the B/T of a galaxy from our SDSS sample with a similar  $M_*$  and sSFR. The intuitive interpretation of this modeling choice is that we assume the morphology-density relation is entirely determined by the color-density relation.

### 2.4 Mock Measurements of Clustering and Lensing

For mock galaxies, to compute galaxy clustering we employ the distant observer approximation by treating the simulation  $z$ -axis as the line-of-sight. We compute  $w_p$  using the `mock_observables.wp` function in **Halotools**, which is a python implementation of the algorithm in the **Corrfunc** C library (Sinha & Garrison 2017).

## 3 GALAXY PROFILE MODEL

Motivated by the Kravtsov (2013) results, we model the half-mass radii of galaxies as separate power law functions of halo virial radius:

$$R_{1/2}^{\text{bulge}} = A_{\text{bulge}} R_{\text{vir}}^{\alpha_{\text{bulge}}} \quad (1)$$

$$R_{1/2}^{\text{disk}} = A_{\text{disk}} R_{\text{vir}}^{\alpha_{\text{disk}}} \quad (2)$$

For central galaxies residing in host halos, we use the present-day virial radius; for satellite galaxies we use the virial radius at the time of accretion, using physical units of kpc. We model the composite size of model galaxies as

$$R_{1/2} = (\text{B/T}) R_{1/2}^{\text{bulge}} + (1 - \text{B/T}) R_{1/2}^{\text{disk}}. \quad (3)$$

When generating Monte Carlo realizations of our model galaxy population, we add log-normal scatter  $\sigma_{R_{1/2}}$  separately to each component. Our model for sizes thus has five free parameters:  $A_{\text{bulge}}, A_{\text{disk}}, \alpha_{\text{bulge}}, \alpha_{\text{disk}},$  and  $\sigma_{R_{1/2}}$ .

We reiterate an important distinction drawn in the formulation of our forward modeling approach. In the model described above,  $R_{1/2}^{\text{bulge}}$  and  $R_{1/2}^{\text{disk}}$  are forward modeling quantities that we aim to constrain through SDSS observations of *composite* galaxy size  $R_{1/2}$ . In particular, both model and real galaxies will be categorized as “small” or “large” according to whether  $R_{1/2}$  is above or below the median size for a galaxy with total stellar mass  $M_*$ .

## 4 RESULTS

In § 4.1 we show comparisons between the fiducial galaxy profile model described in § 3 and our SDSS sample. We demonstrate the sensitivity of galaxy clustering measurements to the post-infall evolution of satellite profiles in § 4.2, establishing the success of our fiducial model as non-trivial. In § 4.3 we show that even highly idealized measurements of galaxy clustering and lensing are insensitive to assumptions about the relation between halo spin  $\lambda_{\text{halo}}$  and size  $R_{1/2}$  of disk-dominated galaxies, implying that large-scale structure observations yield little-to-no constraining power on such assumptions.

### 4.1 Tests of the Fiducial Power Law Model

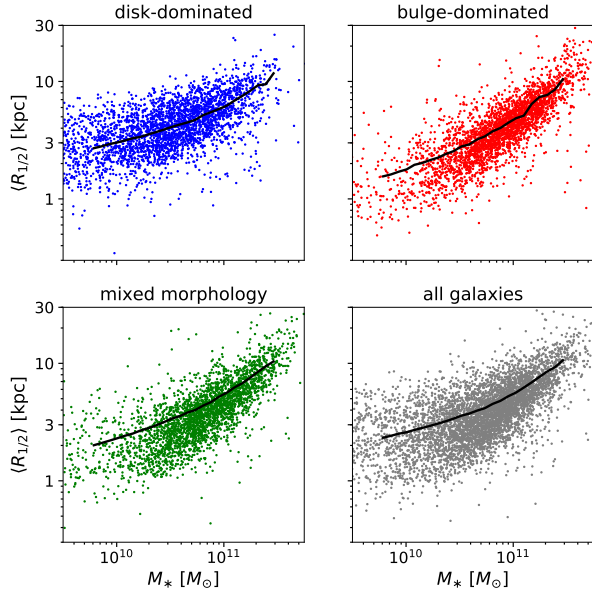
#### 4.1.1 Scaling of $R_{1/2}$ with $M_*$

In Figure 1 we show the scaling of composite galaxy size  $R_{1/2}$  with  $M_*$ , with separate panels dedicated to subsamples of galaxies divided according to value of B/T. We refer to galaxies with  $\text{B/T} < 0.25$  as “disk-dominated”,  $\text{B/T} \geq 0.75$  as “bulge-dominated”, and  $0.25 \leq \text{B/T} < 0.75$  as “mixed morphology”.

The results shown in Kravtsov (2013) simplify our task of parameter space exploration: the assumption that composite galaxy size  $R_{1/2}$  scales in linear proportion to halo radius  $R_{\text{vir}}$  gives a qualitatively good description of SDSS data over many orders of magnitude in galaxy mass. As first pointed out in Kravtsov (2013), this linearization is non-trivial because of the well-known curvature in the scaling relation of  $\langle R_{1/2} | M_* \rangle$ , which is visually apparent in all panels of Fig. 1.

Comparison of the top panels of Fig. 1 show that this linearization is only approximate; evidently, the composite size  $R_{1/2}$  of bulge-dominated galaxies has a steeper scaling relation with  $M_*$  relative to disk-dominated galaxies. This difference motivates our simple extension of the Kravtsov (2013) model, in which composite size  $R_{1/2}$  of disk- and bulge-dominated galaxies scale according to separate power laws.

Through MCMC parameter space exploration aided by **emcee** (Foreman-Mackey et al. 2013), we find a good description of the observed one-point scaling relations is given by model galaxies with the parameters  $A_{\text{disk}} =$



**Figure 1. One-point data used to fit the fiducial model.** Scattered points show the  $R_{1/2} - M_*$  relation for SDSS galaxies as measured in Meert et al. (2015). Bulge-dominated galaxies are defined in terms of the bulge-to-total stellar mass ratio  $B/T \geq 0.75$ , disk-dominated galaxies  $B/T < 0.25$ , mixed morphology as  $0.25 < B/T < 0.75$ . The black curve in each panel shows the  $R_{1/2} - M_*$  relation implied by our fiducial model, in which  $R_{1/2} = AR_{\text{vir}}^\alpha$ , with  $A_{\text{disk}} = 0.014 = 7A_{\text{bulge}}$ ,  $\alpha_{\text{disk}} = 1$ ,  $\alpha_{\text{bulge}} = 5/4$ , and uncorrelated log-normal scatter of 0.2 dex about these relations.

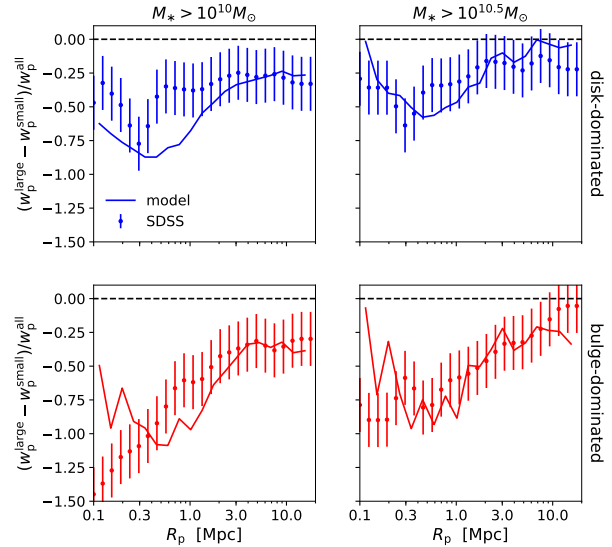
$0.014 = 7A_{\text{bulge}}$ ,  $\alpha_{\text{disk}} = 1$ ,  $\alpha_{\text{bulge}} = 5/4$ , with log-normal scatter of 0.2 dex motivated by Somerville et al. (2017). The black curves in Fig. 1 illustrate the median  $\langle R_{1/2} | M_* \rangle$ .

We confirm the visual impression of comparing the top panels of Fig. 1: we are unable to find a parameter combination in which the composite size of both bulge- and disk-dominated galaxies have linear ( $\alpha = 1$ ) scaling with  $R_{\text{vir}}$ . We do not quantitatively rule out such a scaling relation, as this would require careful investigation of the systematic error budget of galaxy profile modeling that we consider beyond the scope, though well-motivated, by the present work.

#### 4.1.2 Dependence of clustering on $R_{1/2}$

In this section we present new measurements of the  $R_{1/2}$ -dependence of projected galaxy clustering,  $w_p(r_p)$ . Galaxy clustering has complex, simultaneous dependence upon  $M_*$ , sSFR,  $R_{1/2}$ , and  $B/T$ . Since the purpose of this work is to test models of composite galaxy size  $R_{1/2}$ , we design a measurement that is specifically tailored to isolate the influence of  $R_{1/2}$  on the two-point function, while minimizing the dependence upon the other correlated variables.

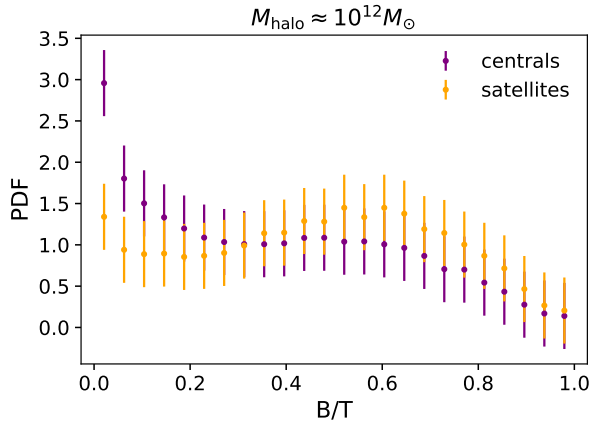
To achieve this, we compute projected clustering separately for  $M_*$ -complete samples of galaxies defined by simultaneous cuts on  $B/T$  and  $R_{1/2}$ , described as follows. First, we construct subsamples of disk-dominated and bulge-dominated galaxies using cuts of  $B/T < 0.25$



**Figure 2. Two-point clustering predictions of the fiducial model.** Points with error bars show new SDSS measurements of the  $R_{1/2}$ -dependence of projected galaxy clustering,  $w_p$ , compared to predictions by the model tuned to the measurements shown in Fig. 1. We define a disk or bulge as “large” or “small” according to whether it is above or below the median size for its stellar mass. The y-axis shows clustering strength ratios, so that, for example, a y-axis value of  $-0.5$  corresponds to small galaxies being 50% more strongly clustered than large galaxies of comparable stellar mass. We show results separately for disk-dominated galaxies (top panels) and bulge-dominated galaxies (bottom panels), and different thresholds in total stellar mass in the left and right panels. The successful prediction shown here is remarkable because the model was not fit to these data, and because two-point clustering is highly sensitive to the physics that shapes satellite galaxy profiles (see Fig. 4).

and  $B/T > 0.75$ , respectively, as in § 4.1.1 above. Separately for each  $B/T$ -selected subsample, we calculate  $\langle R_{1/2} | M_*, B/T \rangle$ , the median composite size of the sample as a function of total galaxy stellar mass. We split each  $B/T$ -selected sample in half according to whether the galaxy has a composite size  $R_{1/2}$  that is above or below this median value. For any  $M_*$ -complete threshold, this gives four subsamples of large and small, disk- and bulge-dominated galaxies. For each  $B/T$ -selected sample, we calculate the difference  $w_p^{\text{large}} - w_p^{\text{small}}$ , scaled by  $w_p^{\text{all}}$ , the clustering of the  $B/T$ -selected sample prior to splitting on  $R_{1/2}$ , and refer to this quantity as *the  $R_{1/2}$  clustering ratio*.

The y-axis of each panel in Figure 2 shows these  $R_{1/2}$  clustering ratios, with left and right panels showing measurements for  $M_* > 10^{10} M_\odot$  and  $M_* > 10^{10.5} M_\odot$  thresholds in total stellar mass, and top and bottom panels showing results for disk- and bulge-dominated samples, respectively. Points with error bars show SDSS measurements, solid curves show the clustering ratios of model galaxies. Before unpacking the information contained in these clustering measurements, we stress that the good agreement shown between model and data in Figure 2 is a genuine model prediction, since we only tuned parameters to fit the one-point measurements in Fig. 1.



**Figure 3. Morphologies of centrals and satellites of the same halo mass.** We show the PDF of the bulge-to-total stellar mass ratio  $B/T$  of central and satellite galaxies of the same halo mass  $M_{\text{halo}} = M_{\text{peak}}$ . In the model, the PDF of  $B/T$  is determined by the stellar mass  $M_*$  and specific star-formation rate  $sSFR$  statistically determined, with no residual dependence on the cosmic web. Satellite galaxies in the model are “bulgier” than centrals of the same halo mass because satellites are more quiescent than centrals of the same stellar mass. Thus our baseline model ansatz is that the morphology-density relation is purely derived from the color-density relation. Since composite size  $R_{1/2}$  in the model is given by  $R_{1/2} = (B/T)R_{1/2}^{\text{bulge}} + (1 - B/T)R_{1/2}^{\text{disk}}$ , this ansatz in turn implies that smaller galaxies cluster more strongly relative to larger galaxies of the same stellar mass, as seen in the measurements shown in Fig. 2.

The salient feature of the clustering ratio measurements is that they are negative: for either disk- or bulge-dominated samples, small galaxies cluster more strongly than large galaxies of the same stellar mass threshold, a new result. This feature also holds true for model galaxies, for which both  $R_{1/2}^{\text{bulge}}$  and  $R_{1/2}^{\text{disk}}$  scale in power law proportion to  $R_{\text{vir}}$ . Since halo mass  $M_{\text{vir}} \propto R_{\text{vir}}^3$ , and clustering is a strong function of  $M_{\text{vir}}$ , the result may seem surprising.

The resolution to this puzzle is illustrated in Figure 3, which shows the PDF of  $B/T$  for halos of the same mass  $M_{\text{halo}}$ , defined as  $M_{\text{peak}}$ , the peak mass ever attained through the history of the (sub)halo, so that host halos and subhalos can be treated on equal footing. In the model, central galaxies are diskier than satellites, a feature that is inherited from the method we use to map  $B/T$  to model galaxies.

In the `UniverseMachine`, and in any quantitatively successful cosmological model of galaxy formation, central galaxies are preferentially more star-forming relative to satellites of comparable stellar or halo mass. Since bluer galaxies are diskier than redder galaxies, our model centrals will be diskier than satellites. Our galaxy profile model has  $A_{\text{disk}} = 7A_{\text{bulge}}$ , and so satellite galaxies will be smaller than centrals due to the  $B/T - sSFR$  correlation. These simultaneous features alone will naturally result in the negative sign of the clustering ratios seen in Figure 2.

In light of this explanation, it is natural to ask

whether this simple feature *alone* is all that is needed for any model of  $R_{1/2}$  to achieve this level of success at predicting galaxy clustering on small and large scales. We address this question in § 4.2 below, finding that the answer is no: the reasonably correct magnitude and scale-dependence of the clustering ratios predicted by our model is a non-trivial result that places tight constraints on the post-infall physics of satellite galaxies.

## 4.2 Sensitivity of Clustering to Satellite Profile Evolution

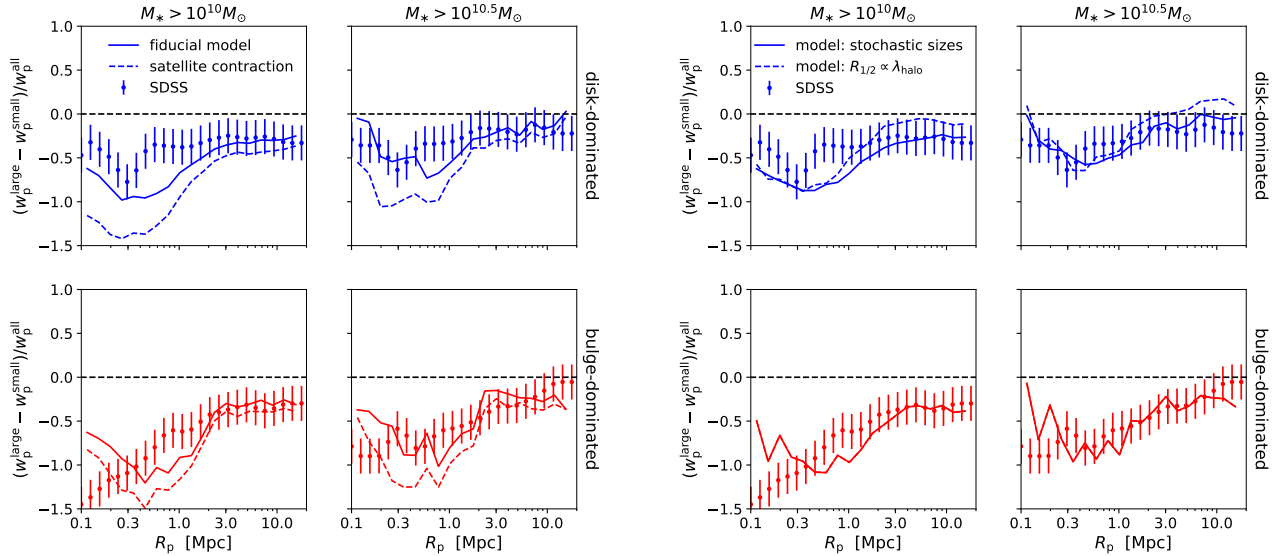
In this section we explore how post-infall changes in satellite  $R_{1/2}$  manifest in  $w_p(r_p)$ . To do so, we construct an extension our fiducial model with a simple additional ingredient for post-infall size evolution of satellites.

In the fiducial model described in § 3, recall that in the power law scaling relations (Eq. 1) the halo radius  $R_{\text{vir}}$  used for satellite galaxies is taken to be the value at the time of infall. The simple interpretation of this assumption is that, in a statistical sense, the size of satellite galaxies is determined by its history as a central galaxy, and that post-infall physics leaves no distinct imprint on satellite  $R_{1/2}$ .

Here we suppose that this assumption is violated, and that  $R_{1/2}$  decreases after infall according to  $(M_{\text{vir}}/M_{\text{acc}})^{1/3}$ . That is, in this alternative formulation, rather than using  $R_{\text{vir}}$  at the time of infall for satellites in the power law scaling relations, we instead use  $R_{\text{vir}}(M_{\text{vir}}/M_{\text{acc}})^{1/3}$ . This simple toy model attempts supposes that the same physical processes leading to halo mass loss are also responsible for a post-infall decrease in satellite size.

Figure 4 compares the clustering ratios of our fiducial model (solid curves) to the clustering ratios of this alternative model (dashed curves). The difference between the small-scale clustering of the two models is stark: the assumption that  $R_{1/2}$  decreases after infall leaves a strong imprint on the  $R_{1/2}$ -dependence of  $w_p(r_p)$ , such that small galaxies cluster *much* more strongly relative to large galaxies, particularly for disk-dominated systems with  $M_* \approx 10^{10} M_{\odot}$ .

The large differences between the solid and dashed curves in Fig. 4 establishes that the largely successful prediction for the clustering ratios fiducial model is a non-trivial:  $w_p(r_p)$  is indeed providing good constraining power on the assumptions underlying our profile modeling and not simply our morphology modeling, c.f., §2.3.2 and §3. We refer the reader to §5.1 for further discussion of the physical implications of this result.



**Figure 4. Clustering tightly constrains satellite-specific size evolution.** Here we compare our fiducial model, in which satellite galaxy size is set by  $R_{\text{vir}}$  at the time of infall, to an alternative model analogous to Watson et al. (2012) in which satellite sizes contract in proportion to  $(M_{\text{vir}}/M_{\text{acc}})^{1/3}$ . The large differences between solid and dashed curves in the top panels show that the  $R_{1/2}$ -dependence of galaxy clustering ratios is highly sensitive to the post-infall evolution of satellite galaxy profiles. The successful prediction of our fiducial model, in which satellite galaxies neither contract nor puff up after infall, places tight constraints on satellite-specific physical processes, which must be either negligible or conspiratorially produce little-to-no size change after accretion.

#### 4.3 Insensitivity of Clustering and Lensing to Halo Spin Correlations

### 5 DISCUSSION

#### 5.1 Physical Interpretation for Satellites

#### 5.2 Progression from Backwards to Forwards Modeling

#### 5.3 Relation to Previous Work

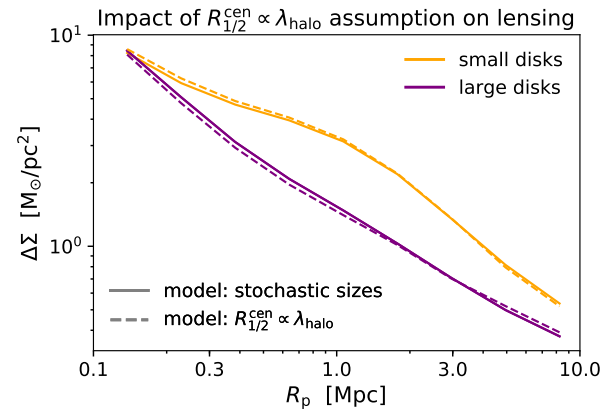
### 6 CONCLUSION

#### 6.1 Summary

#### ACKNOWLEDGMENTS

#### REFERENCES

- Ahn C. P., Alexandroff R., Allende Prieto C., Anders F., Anderson S. F., Anderton T., Andrews B. H., Aubourg É., Bailey S., Bastien F. A., et al. 2014, *ApJS*, 211, 17  
 Behroozi P., Wechsler R., Hearin A., Conroy C., 2017, in prep.  
 Behroozi P. S., Zhu G., Ferguson H. C., Hearin A. P., Lotz J., Silk J., Kassin S., Lu Y., Croton D., Somerville R. S., Watson D. F., 2015, *MNRAS*, 450, 1546  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151



**Figure 5. Clustering and lensing provide no constraining power on the assumption that  $R_{1/2}^{\text{disk}} \propto \lambda_{\text{halo}}$ .** We compare the predictions between our fiducial model in which sizes are purely stochastic, and an alternative model motivated by Mo et al. (1998) in which central galaxy disk size is maximally correlated with host halo spin at fixed stellar mass (implemented via conditional abundance matching, e.g., Hearin et al. 2014). The tiny differences between the solid and dashed curves imply that conventional large-scale structure measurements cannot even in principle provide compelling evidence pertaining to the assumption that  $R_{1/2}^{\text{disk}} \propto \lambda_{\text{halo}}$ .

- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *ASP*, 125, 306  
 Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, *MNRAS*, 444, 729  
 Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, *MNRAS*, 341, 33  
 Kravtsov A. V., 2013, *ApJL*, 764, L31  
 Meert A., Vikram V., Bernardi M., 2015, *MNRAS*, 446, 3943  
 Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319  
 Sinha M., Garrison L., 2017, *Corrfunc: Blazing fast correlation functions on the CPU*, Astrophysics Source Code Library  
 Somerville R. S., Behroozi P., Pandya V., et al., 2017,

ArXiv e-prints

Watson D. F., Berlind A. A., Zentner A. R., 2012, ApJ ,  
754, 90