MNRAS 000, 1–?? (2023) Preprint 28 August 2023 Compiled using MNRAS LATEX style file v3.0

Breaking degeneracies in the first galaxies with clustering

Julian B. Muñoz,[1⋆] Jordan Mirocha,[2,3] Steven Furlanetto,[4] and Nashwan Sabti[5]
[1] Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA
[2] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[3] California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA
[4] Department of Physics & Astronomy, University of California, Los Angeles, Los Angeles, CA 90095, USA
[5] William H. Miller III Department of Physics and Astronomy, 3400 N. Charles St., Baltimore, MD 21218, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The high-redshift galaxy UV luminosity function (UVLF) has become essential for understanding the formation and evolution of the first galaxies. Yet, UVLFs only measure galaxy abundances, giving rise to a degeneracy between the mean galaxy luminosity and its stochasticity. Here, we show that upcoming clustering measurements with the James Webb Space Telescope (JWST), as well as with Roman, will be able to break this degeneracy, even at redshifts $z \gtrsim 10$. First, we demonstrate that current Subaru Hyper Suprime-Cam (HSC) measurements of the galaxy bias at $z \sim 4-6$ point to a relatively tight halo-galaxy connection, with low stochasticity. Then, we show that the larger UVLFs observed by JWST at $z \gtrsim 10$ can be explained with either a boosted average UV emission or an enhanced stochasticity. These two models, however, predict different galaxy biases, which are potentially distinguishable in JWST and Roman surveys. Galaxy-clustering measurements, therefore, will provide crucial insights into the connection between the first galaxies and their dark-matter halos, and identify the root cause of the enhanced abundance of $z \gtrsim 10$ galaxies revealed with JWST during its first year of operations.

Key words: galaxies: high-redshift – dark ages, reionization, first stars – cosmology: theory – intergalactic medium – diffuse radiation

1 INTRODUCTION

The ultraviolet luminosity function (UVLF) has emerged as a crucial observational tool to understand the formation of the first galaxies during cosmic dawn and reionization (e.g., Finkelstein 2016). Analyses of the UVLF have derived important constraints on the evolution of high-redshift galaxies (Trenti et al. 2010; Ceverino et al. 2017; Tacchella et al. 2018) and the process of cosmic reionization (Mirocha et al. 2017; Park et al. 2019; Mason et al. 2019), as well as set new cosmological bounds (Corasaniti et al. 2017; Menci et al. 2018; Sabti et al. 2021; Rudakovskyi et al. 2021; Sabti et al. 2022b; Sabti et al. 2023b). In practice, fitting the UVLFs has become a litmus test for any model of high-z galaxy formation.

Deep imaging surveys with space-based telescopes, such as the *Hubble* and *James Webb* Space Telescopes (HST and JWST, respectively), have extended the reach of UVLFs to ever-increasing redshifts, pushing the observational frontiers closer to the cosmic dawn (Livermore et al. 2017; Atek et al. 2015; Bouwens et al. 2015; Finkelstein et al. 2022a; Treu et al. 2022; Eisenstein et al. 2023). Intriguingly, the abundance of star-forming galaxies at $z \gtrsim 10$, reported from the first JWST observations in Finkelstein et al. (2022b); Harikane et al. (2023b); Donnan et al. (2023); Bouwens et al. (2023); Finkelstein et al. (2022a); Castellano et al. (2022); Naidu et al. (2022); Leung et al. (2023); Adams et al. (2023), appears at odds with expectations: there seems to be significantly more UV-bright galaxies than predicted by most galaxy-formation models. While only a subset of candidates have been spectroscopically confirmed so far (Fujimoto et al. 2023; Arrabal Haro et al. 2023; Curtis-Lake et al. 2022; Harikane et al. 2023a; Kocevski et al. 2023), with the highest-redshift one identified to be a contaminant (Arrabal Haro et al. 2023; Zavala et al. 2023), there has been a flurry of theoretical activity to explain the photometrically derived UVLFs, with solutions broadly coming in two flavors: either an enhanced star-formation/UV emission at high z (Ferrara et al. 2022; Inayoshi et al. 2022; Dekel et al. 2023; Yung et al. 2023; Steinhardt et al. 2022), or a large "stochasticity" in the UV brightness of the first galaxies (Mason et al. 2023; Mirocha & Furlanetto 2023; Shen et al. 2023; Padmanabhan & Loeb 2023). Both models are able to explain the JWST data by producing more UV-bright galaxies, in the former case as an average and in the latter as the tail of a distribution.

This exemplifies an *intrinsic degeneracy* in the UVLFs. UVLFs measure the abundance of galaxies, which would would be sufficient for high-z studies if galaxy luminosities were linked one-to-one to their host dark-matter halo prop-

⋆ E-mail: julianbmunoz@utexas.edu

erties (Behroozi et al. 2010; Moster et al. 2013). However, a more complex—or stochastic—halo-galaxy connection gives rise to more than one way to obtain the same average density of galaxies (Ren et al. 2019; Mirocha 2020). We illustrate this degeneracy in Fig. 1, where we show an array of models that can explain the observed UVLF at $z \sim 4$ (from Bouwens et al. 2021), but have vastly different halo-galaxy connections (with stronger or weaker stochasticity).

Here we argue that galaxy clustering can break this degeneracy at high redshifts, akin to halo-occupation-distribution (HOD) studies at lower z (e.g., Giavalisco & Dickinson 2001; Kravtsov et al. 2004; Zheng & Weinberg 2007; Moster et al. 2010; Zentner et al. 2019; Hadzhiyska et al. 2021), and that this clustering can be measured by current-generation surveys. Galaxies residing in heavier dark-matter halos will cluster more strongly than their lighter counterparts, boasting a larger galaxy bias. As such, measuring clustering at high z will allow us to determine whether UV-bright galaxies are rare because their host halos are rare as well, or because they only form stars efficiently some fraction of the time. To illustrate this point, we will show how current bias data at $z \sim 4-6$ tend to prefer a closer halo-galaxy connection (with lower stochasticity, and a duty cycle near unity). Turning to $z \gtrsim 10$, we will argue that upcoming surveys, like Cosmos-Web with JWST (Casey et al. 2022) and the *Roman* high-latitude survey (HLS, Spergel et al. 2015), will be able to distinguish between models of galaxy formation. In particular, they will help determine whether the first galaxies showed enhanced stochasticity or higher UV emission. Throughout this work, we will fix our cosmological parameters to $h = 0.7$ and $\Omega_{\rm m} = 0.3$ to match Bouwens et al. (2021), set $\sigma_8 = 0.85$, $n_{\rm s} = 0.966$, $\omega_{\rm b} = 0.022$, and use AB magnitudes (Oke & Gunn 1983).

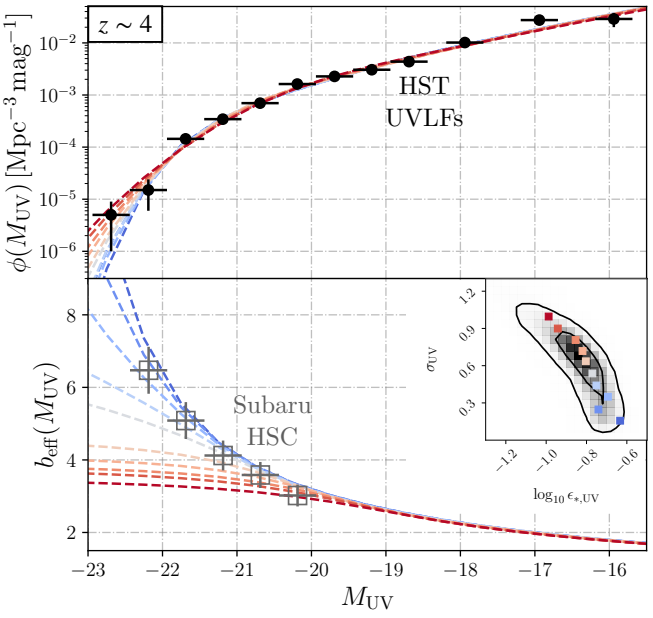

Figure 1. Illustration of the degeneracy between stochasticity and star-formation efficiency at high z. We show the UVLF (**top**) and the bias (**bottom**) at $z \sim 4$ as a function of UV magnitude for different models. These are all calibrated to the UVLF data (black circles, from Bouwens et al. 2021) and colored by their standard deviation $\sigma_{\rm UV}$ (in the $P(M_{\rm UV}|M_{\rm h})$ distribution), whose posterior against the star-formation efficiency $\epsilon_{\star,{\rm UV}}$ is shown in the inset. The bias predictions differ dramatically, depending on whether heavy or light halo masses are responsible for the bright end of the UVLF. We show the biases reported in Harikane et al. (2022) as gray open squares, which have not been used in the inference. All predictions are binned with $\Delta M_{\rm UV} = 0.5$.

2 THE DEGENERACY IN THE UVLFS

Let us begin by laying out our definitions, and with them the intrinsic degeneracy in the UVLF. The UVLF measures the number density of objects with a particular UV magnitude $M_{\rm UV}$. Assuming all galaxies live in dark-matter halos (and a halo-occupation fraction of unity), we can find the UVLF as:

$$\phi_{\rm UV} \equiv \frac{{\rm d}n}{{\rm d}M_{\rm UV}} = \int {\rm d}M_{\rm h} \frac{{\rm d}n}{{\rm d}M_{\rm h}} P(M_{\rm UV}|M_{\rm h}) \ , \quad (1)$$

where ${\rm d}n/{\rm d}M_{\rm h}$ is the halo mass function, following the Sheth & Tormen (2002) fit (see also Rodriguez-Puebla et al. 2016), and $P(M_{\rm UV}|M_{\rm h})$ is the probability that a halo of mass $M_{\rm h}$ hosts a galaxy with magnitude $M_{\rm UV}$. The UVLF is often reported in bins:

$$\phi^{\rm bin}_{\rm UV} = \int {\rm d}z\, W_z(z) \int {\rm d}M_{\rm UV} W_{\rm UV}(M_{\rm UV}) \phi_{\rm UV}(M_{\rm UV}, z) \ , \quad (2)$$

where both window functions W_i integrate to unity. The magnitude windows $W_{\rm UV}$ are assumed to be tophats, whereas the W_z are Gaussians fit to the redshift distributions in Bouwens et al. (2021, where it is important for data and predictions to match the true mean and width z, see also Trapp et al. 2022).

The $P(M_{\rm UV}|M_{\rm h})$ term, despite its apparent simplicity, encodes the complex halo-galaxy connection (Wechsler & Tinker 2018), including any stochasticity. We will model it through a semi-analytic approach where $P(M_{\rm UV}|M_{\rm h})$ is a Gaussian centered around a predicted "mean" $\overline{M_{\rm UV}}(M_{\rm h})$, with a mass-independent dispersion $\sigma_{\rm UV}$ that we allow to vary. The UV magnitude of a galaxy will depend on its star-formation rate (SFR):

$$\dot{M}_\star = f_\star(z, M_{\rm h}) f_{\rm b} \dot{M}_{\rm h} \ , \quad (3)$$

where $f_{\rm b} \approx 0.16$ is the baryon fraction (Aghanim et al. 2020), and we take a model of exponential accretion $M_{\rm h}(z) \propto e^{a_{\rm acc} z}$ with $a_{\rm acc} = 0.79$ (Schneider et al. 2021, see also App. A). The average halo-galaxy connection is encoded in the star-formation efficiency (SFE) $f_\star$, generically a function of redshift z and halo mass $M_{\rm h}$. Inspired by analytic and simulation studies (Moster et al. 2010; Furlanetto et al. 2017), we assume a double-power law functional form:

$$f_\star = \frac{2\epsilon_\star}{(M_{\rm h}/M_c)^{-\alpha_\star} + (M_{\rm h}/M_c)^{-\beta_\star}} \ , \quad (4)$$

which was shown to fit well both observations and simulations in Sabti et al. (2022a). This $f_\star$ has four free parameters: an amplitude $\epsilon_\star$, a critical mass M_c, and two power-law indices $\alpha_\star > 0$ and $\beta_\star < 0$ for the faint and bright ends, respectively. The SFR is converted to UV luminosity through $L_{\rm UV} = \dot{M}_\star/\kappa_{\rm UV}$. Given that the conversion factor $\kappa_{\rm UV}$ is fully degenerate with $\epsilon_\star$, we will define:

$$\epsilon_{\star,{\rm UV}} \equiv \epsilon_\star \left(\kappa_{\rm UV}/\overline{\kappa_{\rm UV}}\right)^{-1} \ , \quad (5)$$

for a fiducial $\overline{\kappa_{\rm UV}} = 1.15 \times 10^{-28}$ $(M_\odot\,{\rm yr}^{-1})$ /(erg s^{-1}) as in Madau & Dickinson (2014). The UV luminosity per unit SFR could be higher for a top-heavy initial mass function (as expected for Population III stars, Bromm et al. 2002), and varying $\epsilon_{\star,\rm UV}$ will allow us to capture such behavior. In addition, we ought to account for dust attenuation, which reduces the apparent brightness of galaxies, especially towards lower z and the bright end (Yung et al. 2018). Following Sabti et al. (2022a), we adopt the IRX-β dust prescription calibrated by Meurer et al. (1999) with the β measured in Bouwens et al. (2014). We extrapolate to $z > 8$ by using the $z = 8$ result, as advocated in Mason et al. (2023).

While simple, this semi-analytical formalism is highly predictive, though it is not without caveats. One has to allow the free parameters to vary with z to capture the time-dependent impact of feedback, as well as float $\sigma_{\rm UV}$ against $M_{\rm UV}$ to study its luminosity dependence. In this first work we will account for the former effect only. We fit the five free parameters (four in the $f_\star$ relation plus $\sigma_{\rm UV}$) to the $z = 4$ UVLF data from Bouwens et al. (2021)[1], and show in Fig. 1 the results for an array of parameters within the 2σ preferred region, colored by their stochasticity $\sigma_{\rm UV}$. From this figure it is clear that there is a degeneracy between the *average* halo-galaxy connection $f_\star$ and its stochasticity $\sigma_{\rm UV}$, most obvious in the $\epsilon_{\star,\rm UV} - \sigma_{\rm UV}$ plane. In essence, there are two ways to explain the paucity of objects in the bright end of the UVLF. One option is that those galaxies reside in very massive (and therefore rare) halos in the early universe. The other is that these galaxies are hosted in lower-mass halos, but correspond to the high-brightness tail of a stochastic halo-galaxy connection. This degeneracy is intrinsic to one-point functions, like the UVLFs.

Here we argue that these two scenarios (which can be rephrased in terms of a duty cycle $f_{\rm duty}$, see Lee et al. 2006 and our App. B), are distinguishable by using clustering information, such as two-point functions, as massive halos cluster more strongly than lighter ones. In this work, we will examine clustering via the halo bias $b(M_{\rm h})$, the linear ratio of the galaxy and matter densities (Desjacques et al. 2018, though of course more information can be gleaned through nonlinear studies; Berlind & Weinberg 2002; Cooray & Sheth 2002), fitted in Tinker et al. (2010). Using our formalism, we calculate the number-weighted effective bias as:

$$b_{\rm eff}(M_{\rm UV}) = \phi_{\rm UV}^{-1} \int {\rm d}M_{\rm h} \frac{{\rm d}n}{{\rm d}M_{\rm h}} b(M_{\rm h}) P(M_{\rm UV}|M_{\rm h}) \ , \tag{6}$$

which we bin as in Eq. (2) and show for each of our models in the bottom panel of Fig. 1. Models with large stochasticity ($\sigma_{\rm UV}$) predict a very mild dependence of bias against intrinsic magnitude, as the bright end will be dominated by smaller-mass objects that have been upscattered in luminosity. Conversely, a tighter halo-galaxy connection (lower $\sigma_{\rm UV}$) predicts a sharp rise in the bias towards the bright end, as those objects reside in more massive (and thus exponentially suppressed) halos. The behavior of $b_{\rm eff}$ is therefore critical to break degeneracies in the halo-galaxy connection, especially towards the bright end.

Along with our predictions, Fig. 1 shows bias measurements as reported in Harikane et al. (2022), which were obtained from clustering studies with the *Subaru* Hyper Suprime-Cam (HSC) at $z = 4$. While adding these measurements to our likelihood is tempting, a word of caution is warranted. The HSC biases were inferred for several magnitude cuts using an HOD model fitted within a different fiducial cosmology than ours. As such, we converted these measurements for illustration purposes only (see App. C for our procedure), and show them with a 10% noise floor to account for these uncertainties, leaving a direct likelihood analysis to future work. Nevertheless, it is clear that the bias measured with HSC rises sharply towards brighter magnitudes, which indicates a preference for lower stochasticity ($\sigma_{\rm UV} \lesssim 0.5$ in our model). In Harikane et al. (2022), this appeared as a close connection in the $M_{\rm h} - M_{\rm UV}$ plane (e.g., their Fig. 16). We refer the interested reader to App. D, where we repeat this analysis by simultaneously varying $\sigma_{\rm UV}$ and $\epsilon_{\star,\rm UV}$ at each redshift slice in the $z = 5 - 7$ range, finding similar results.

It is then clear that galaxy clustering has the potential to break the degeneracy between the average halo-galaxy connection ($f_\star$) and its stochasticity ($\sigma_{\rm UV}$). Let us now study the impact of galaxy clustering in light of the $z \gtrsim 10$ JWST data.

3 FITTING THE HIGH-REDSHIFT UVLFS

So far, we have shown that at a single redshift there is an intrinsic degeneracy in the UVLFs between the amplitude of the SFE of galaxies and their stochasticity. Of course, if these two parameters had a known z evolution, co-adding information from different redshifts could help break this degeneracy. There is, however, no guarantee that either $\epsilon_{\star,\rm UV}$ or $\sigma_{\rm UV}$ are constant with z. For instance, the shape of the SFE $f_\star$ can evolve with the strength of astrophysical feedback (Furlanetto et al. 2017). Even with a fixed SFE, galaxies become more metal-rich over time, which affects the $\kappa_{\rm UV}$ conversion between SFR and UV emission (and thus $\epsilon_{\star,\rm UV}$). Likewise, $\sigma_{\rm UV}$ could evolve as reionization and feedback strip gas away from small-mass galaxies (Faucher-Giguère 2018; Furlanetto & Mirocha 2022b). Rather than fixing these parameters, we will use two simple models to determine their behavior as a function of redshift. In the first, we will vary $\sigma_{\rm UV}$ independently at each redshift, and vary the rest of parameters linearly with z (specifically, the two power-law indices $\alpha_\star$ and $\beta_\star$, as well as $\log_{10} M_c$ and $\log_{10} \epsilon_{\star,\rm UV}$). In the second model, we will switch the roles of $\log_{10} \epsilon_{\star,\rm UV}$ and $\sigma_{\rm UV}$, making the latter linear in z and the former vary at each slice.

We begin by fitting the HST UVLFs compiled in Bouwens et al. (2021). These are derived from blank-field surveys in the Legacy Fields catalog covering $z = 4 - 10$ and do not include lensing fields. We will fit both models, as our benchmarks, with a (log) likelihood at each redshift:

$$-\log \mathcal{L}(z) = \sum_i \frac{(\phi_{{\rm UV},i}^{\rm obs} - \phi_{{\rm UV},i}^{\rm bin})^2}{2(\sigma_i^{\rm obs})^2} \ , \tag{7}$$

given UVLF measurements $\phi_{{\rm UV},i}^{\rm obs}$ with uncorrelated Gaussian errors $\sigma_i^{\rm obs}$, and our predicted $\phi_{{\rm UV},i}^{\rm bin}$, where the sum runs over magnitude bins i. We then sum the log-likelihoods for all the z we consider (implicitly assuming they are uncorrelated as well). We run a Markov Chain Monte Carlo for both models and find good fits in each case, with $\chi^2 = 33.9$ and 34.6 for

[1] To account for cosmic variance, we impose a minimum 20% error in the UVLF data (Sabti et al. 2022a).

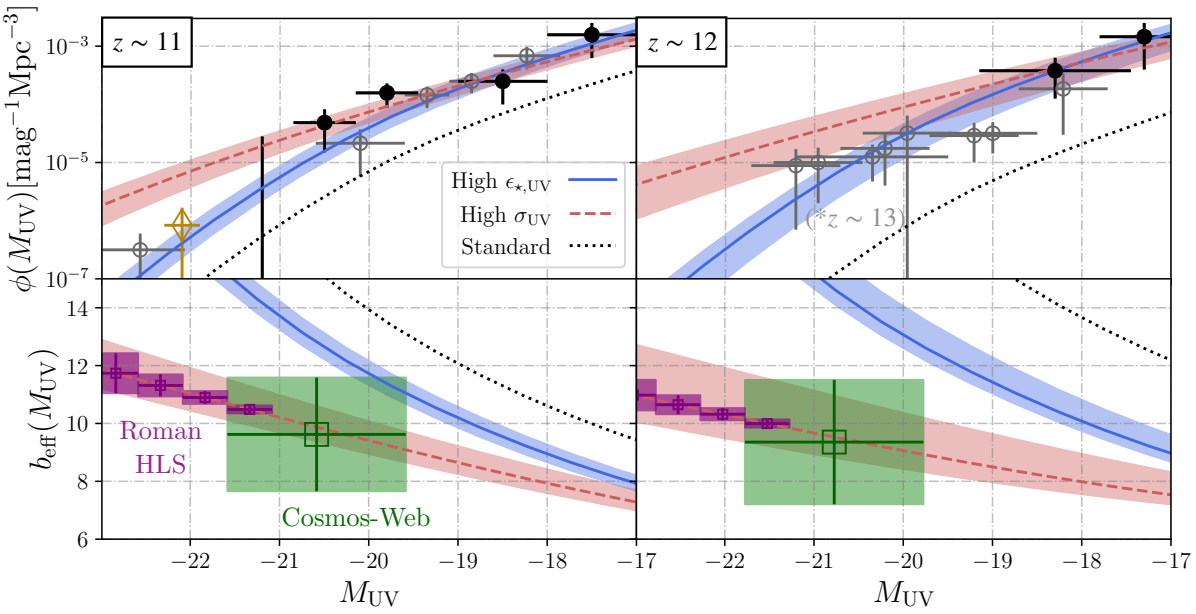

Figure 2. UVLFs and biases at $z \sim 11$ ($z = 10.75$ with a 0.5 rms) and 12 ($z = 12.25$ and 0.75 rms). Filled black points correspond to the UVLF data used in our analysis, from Finkelstein et al. (2022a); Pérez-González et al. (2023); Harikane et al. (2023b), whereas gray points are data from Donnan et al. (2023); Bouwens et al. (2023), which correspond to $z \sim 10.5$ and $z \sim 13$, as indicated. The bright ($M_{\rm UV} < -21$) end of the UVLF at $z \sim 11$ contains two objects: GN-z11 (as a yellow diamond, Oesch et al. 2016), recently argued to host an AGN (Maiolino et al. 2023), and an object that could be an interloper at $z \sim 3$ (Kauffmann et al. 2022), which we do not use in our inference. We show the 1σ predictions for two models, in red for a enhanced stochasticity $\sigma_{\rm UV}(z)$ and in blue for a larger mean UV emission $\epsilon_{\star,\rm UV}(z)$. The black-dotted line is a simple extrapolation of the $z \leq 10$ fit from HST, which does not reach the values necessary to explain the JWST UVLF. Bottom panels show the bias expected in each model, as well as projected measurements for a Cosmos-Web-like survey and the *Roman* high-latitude survey (HLS).

the variable-$\sigma_{\rm UV}$ and $\epsilon_{\star,\rm UV}$ models, respectively. This is with 46 degrees of freedom, indicating perhaps an overestimation of the error bars due to the 20% minimum error imposed to account for cosmic variance. The $\Delta\chi^2 = 0.7$ difference between the two scenarios shows that both are equally good fits to the HST data.

These two runs, calibrated at $z \leq 10$, are designed to provide a "standard" against which to compare the JWST data. As we will see, a naive extrapolation to $z \geq 10$ greatly underpredicts the amount of galaxies observed in JWST. Let us quantify this statement.

The JWST data are not yet constraining enough on their own, so we will leverage the $z \leq 10$ information from HST. In order to minimize covariances and avoid double-counting, we will only include JWST galaxy candidates above $z = 10$. In particular, we build the UVLF at $z \sim 11$ and 12 using the results from CEERS (Finkelstein et al. 2022a and Pérez-González et al. 2023, which can be co-added as the latter only covers fainter objects than the former), and at $z \sim 13$ using data from Harikane et al. (2023b). These are shown in Fig. 2, along with the UVLFs from Bouwens et al. (2023); Donnan et al. (2023). We note that we do not use any $z \sim 16$ data as they are currently highly uncertain, and take the appropriate average of asymmetric error bars. Our UVLF data does not contain the two brightest objects reported at $z \sim 11$ (shown as empty symbols in Fig. 2). They correspond to GN-z11 (Oesch et al. 2016), which has recently been argued to host an active galactic nucleus (AGN, Maiolino et al. 2023), and a galaxy previously reported in COSMOS2020 (ID 1356755) that has secondary solutions at $z \sim 3$ (Kauffmann et al. 2022). These uncertainties underscore the necessity for beyond-UVLF measurements to break degeneracies, as the bright tail of the UVLF is susceptible to contamination from interlopers (Furlanetto & Mirocha 2022a) and AGN (Finkelstein & Bagley 2022).

In practice, we first fit both our models to the HST ($z = 4-10$) data, as outlined above, and fix all parameters to the best fit of that analysis. This then allows us to easily vary the parameter of interest (either $\sigma_{\rm UV}$ or $\epsilon_{\star,\rm UV}$) at higher z without altering the good fit to low-z HST data. Fig. 2 shows the UVLF data at $z \sim 11$ and 12, along with the predictions of the two models (varying $\sigma_{\rm UV}$ and $\epsilon_{\star,\rm UV}$ at these z within their 1σ region), both of which provide a good fit to the JWST data. In particular, the χ^2 difference between the models (co-adding $z \sim 11-13$ data) is $\Delta\chi^2 = 2.5$, weakly favoring the $\sigma_{\rm UV}$ model. Both have $\Delta\chi^2 \approx 15$ compared to the standard prediction, obtained by extrapolating from $z \leq 10$ (also shown in Fig. 2). Thus, the $z > 10$ UVLFs appear to prefer a new component in our models, either large stochasticity or UV brightness.

In order to visualize this, we show the results of varying both parameters over the entire range of $z = 4-13$ in Fig. 3 (see App. E for the rest of astrophysical parameters). In both models we find a consistent trend, where the parameter of interest ($\epsilon_{\star,\rm UV}$ or $\sigma_{\rm UV}$) has a fairly smooth, roughly constant behavior for $z \leq 10$, but abruptly rises at $z > 10$ to explain the overabundance of JWST galaxies. In the first case, we require very efficient UV emission with $\epsilon_{\star,\rm UV} \sim O(1)$, which may be due to feedback-free starbursts (Dekel et al. 2023), or a top-heavy stellar initial mass function (Steinhardt et al.

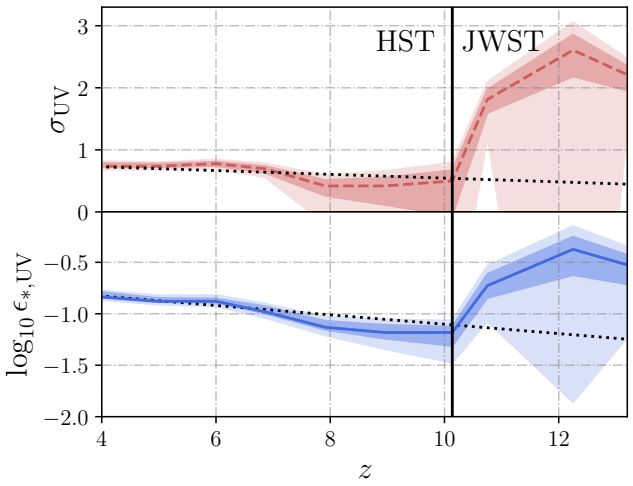

Figure 3. Posteriors for the two models we consider: one with a z-dependent stochasticity $\sigma_{\rm UV}$ (**top**), and one with an enhanced UV emission $\epsilon_{\star,\rm UV}$ (**bottom**). In each panel, we fix the rest of parameters to their best-fit HST values, and show 1σ and 2σ contours in dark and light shades, respectively. Dotted lines correspond to the best-fit linear evolution in z for each model, which is dominated by the (more constraining) low-z data. While the $z \leq 10$ data has a smooth behavior in both variables, the $z = 11-13$ data from JWST prefers a jump towards high stochasticities or UV efficiencies.

2022). In the second case, JWST data demand a very large stochasticity, $\sigma_{\rm UV} \approx 2$, which translates into a full dex of variation in the UV luminosity from galaxy to galaxy, much in excess of what is expected of mass-accretion histories (Mirocha et al. 2021; Mason et al. 2023). Moreover, such large stochasticity may require coherent bursts of star formation lasting $\sim$ tens of Myrs, rather than short, noise-like variations, as spectral synthesis dampens the fluctuations in luminosity. Let us now explore how to distinguish between these two scenarios.

4 CLUSTERING AS A HIGH-REDSHIFT MODEL DISCRIMINATOR

Clustering measurements will provide information on the hosts of the high-z galaxies. We show our predicted biases for both models discussed above in Fig. 2, along with the first JWST UVLFs at $z \sim 11$ and 12. The model with enhanced stochasticity predicts lower biases on average, as the more abundant lighter halos will dominate the bright end of the UVLF. This is true even at magnitudes where the two models show similar UVLFs (e.g., $M_{\rm UV} = -20$ to -17). Thus, while different halo-galaxy connections may give rise to an identical one-point function (UVLF), the higher-order statistics (in our case the bias) will allow us to distinguish between models, even at $z \gtrsim 10$.

While the clustering of the first galaxies may hold the key to understanding their formation, measuring it is no small feat. Studies using HST (Barone-Nugent et al. 2014; Lee et al. 2006; Overzier et al. 2006) and HSC (Harikane et al. 2018, 2022) have quantified clustering up to $z \sim 7$. Extending these measurements to the $z \gtrsim 10$ regime is complicated due to the dearth of high-z galaxies. There is reason, however, to be cautiously optimistic. The JWST-derived UVLFs are much larger than predicted, which means that if they hold to spectroscopic scrutiny we could expect significantly more targets at these z. Moreover, the expected biases are fairly large ($b_{\rm eff} \sim 8-10$, owing to the rarity of nonlinear structures at these high z), easing a detection of the clustering signal. Let us now perform a simple forecast.

We will focus on two survey configurations, the Cosmos-Web JWST survey, which we approximate as covering an area $\Omega = 0.5\,{\rm deg}^2$, up to $m_{\rm UV} = 28.0$ (Casey et al. 2022); and the proposed *Roman* HLS, covering $\Omega = 2200\,{\rm deg}^2$, up to $m_{\rm UV} = 26.5$ (Wang et al. 2022). These represent two near- and mid-future possibilities to measure clustering at high z, though of course other data-sets, including deeper JWST and *Roman* surveys, will add further information. For convenience, we will use a Fisher-matrix formalism as introduced in Jungman et al. (1996):

$$F_{bb} = f_{\rm sky} \sum_{\ell > \ell_{\rm min}} \left(\ell + \frac{1}{2}\right) \left(\frac{\partial C_\ell / \partial b_{\rm eff}}{C_\ell + N_\ell}\right)^2 , \tag{8}$$

with $\ell_{\rm min} = f_{\rm sky}^{-1/2}$ and $f_{\rm sky} = \Omega/(4\pi)$ the sky fraction covered by each survey. We take the noise to be Poissonian, such that $N_\ell = \overline{n}_{\rm 2D}^{-1}$, where

$$\overline{n}_{\rm 2D} = \phi_{\rm UV}^{\rm bin} \Delta M_{\rm UV} \chi^2 \Delta\chi \tag{9}$$

is the surface density of galaxies in each bin, assuming an observational window centered at a comoving distance $\chi(z)$ with a width $\Delta\chi$. As for the observed power spectrum, we take the Limber approximation (LoVerde & Afshordi 2008):

$$C_\ell(z) = \frac{b_{\rm eff}^2}{(\Delta\chi)^2} \int_{\chi-\Delta\chi/2}^{\chi+\Delta\chi/2} \frac{{\rm d}\chi'}{\chi'^2} P_{\rm m}\left[(\ell+1/2)/\chi', z\right] , \tag{10}$$

where we ignore the redshift evolution of $b_{\rm eff}$ and the matter power spectrum $P_{\rm m}$ in the integral.

Under these approximations, we can estimate the expected uncertainty in the bias as $\sigma(b_{\rm eff}) = F_{bb}^{-1/2}$, which we show in Fig. 2 at $z \sim 11$ and 12. At both redshifts, we find that with Cosmos-Web we can expect a $\sim 20\%$ constraint on the bias over a $\Delta M_{\rm UV} = 2.0$ bin, potentially enough to distinguish between our two models and to pinpoint the mass of the halos that hosted such bright galaxies. Moreover, the *Roman* HLS would reduce these errors to the $1-3\%$ level over several magnitude bins, providing a definitive test of our galaxy-formation models.

We note that this forecast was not designed to capture the detailed physics of galaxy clustering at high z, including covariances between data-points (Harikane et al. 2022), light-cone effects (Yung et al. 2022), and the non-linear 1-halo term (Cooray & Sheth 2002, which would boost the signal). Instead, it provides a proof-of-principle that the biases we expect at $z \sim 10$ differ between galaxy-formation models, and are potentially measurable, given the enhancement in the expected number of galaxies from the JWST UVLFs. We leave a detailed analysis for future work (Sabti et al. 2023a), including studying the mass dependence of $\sigma_{\rm UV}$, though we note that previous *Roman* forecasts in Waters et al. (2016); La Plante et al. (2023) have shown that the linear bias is potentially measurable up to $z \sim 13$.

5 DISCUSSION AND CONCLUSIONS

The first JWST-derived UV luminosity functions (UVLFs) at $z > 10$ point to a very active universe, with far more UV-bright galaxies than predicted by our models. The root cause of this discrepancy is, however, difficult to determine from the UVLFs alone. This observable suffers from an intrinsic degeneracy between changes in the average halo-galaxy connection and its stochasticity, as illustrated in Fig. 1. Here we have argued that galaxy-clustering measurements will allow us to break this degeneracy, providing an essential test of galaxy-formation models.

We have performed an analysis based on a flexible, semi-analytic model, which has been tested against HST data and hydrodynamical simulations in Sabti et al. (2022a). We calibrated our model to the $z \leq 10$ HST UVLFs and found that $\sigma_{\rm UV}$ (which accounts for stochasticity) and the star-formation efficiency $f_\star$ are indeed highly degenerate at each z. However, we have shown that clustering measurements from Harikane et al. (2022, using *Subaru* HSC) tentatively point to a tight halo-galaxy connection at $z \sim 4-6$, with $\sigma_{\rm UV} \lesssim 0.5$. This value is in line with the predicted variability in the assembly of dark-matter halos (Mirocha et al. 2021), as well as with previous studies using closest-neighbor clustering (Ren et al. 2018) and field-to-field variations (Robertson 2010).

We have additionally applied our analysis to the first JWST UVLFs. Theoretical work has argued for either an increased stochasticity ($\sigma_{\rm UV}$) or mean UV emission ($\epsilon_{\star,{\rm UV}}$) at $z \gtrsim 10$ to explain the abundance of galaxy candidates in JWST. We have shown that, while both solutions fit well the JWST UVLFs, the large-$\sigma_{\rm UV}$ scenario predicts consistently lower biases. Further, these biases are potentially measurable at $z \gtrsim 10$ with the Cosmos-Web survey, and in the future by *Roman*, allowing us to distinguish between the two galaxy-formation scenarios. Clustering can provide a more robust check than studying the bright-end of the UVLF alone, as the latter is susceptible to systematic effects such as AGN contamination and dust.

In summary, the launch of JWST has ignited a revolution in our understanding of the formation and evolution of the first galaxies. Here we have argued that clustering measurements are critical to break degeneracies in galaxy models. This will allow us to extract the maximum amount of information from JWST and future observatories like *Roman*, and will unveil the connection between the large-scale structure and the first luminous objects in our universe.

Acknowledgements

We are very grateful to Mike Boylan-Kolchin, Caitlin Casey, Daniel Eisenstein, and Steve Finkelstein for helpful discussions, as well as to the anonymous referee for insightful comments. JBM was supported by the UT Board of Visitors, and thanks the ITC at the Harvard-Smithsonian Center for Astrophysics for their hospitality during part of this work. JM was supported by an appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory / California Institute of Technology, administered by Oak Ridge Associated Universities under contract with NASA. SRF was supported by the National Science Foundation through awards AST-1812458 and AST-2205900 and by NASA through award 80NSSC22K0818. NS was supported by a Horizon Fellowship from Johns Hopkins University.

Data Availability

The data underlying this article will be shared on reasonable request to the author.

Software: `numpy` (van der Walt et al. 2011), `emcee` (Foreman-Mackey et al. 2013), `corner` (Foreman-Mackey 2016), `CLASS` (Blas et al. 2011), and `Zeus21` (Muñoz 2023).

REFERENCES

Adams N. J., et al., 2023, MNRAS, 518, 4755
Aghanim N., et al., 2020, Astron. Astrophys., 641, A6
Arrabal Haro P., et al., 2023, arXiv e-prints, p. arXiv:2303.15431
Atek H., et al., 2015, Astrophys. J., 814, 69
Barone-Nugent R. L., et al., 2014, Astrophys. J., 793, 17
Behroozi P. S., Conroy C., Wechsler R. H., 2010, Astrophys. J., 717, 379
Berlind A. A., Weinberg D. H., 2002, Astrophys. J., 575, 587
Blas D., Lesgourgues J., Tram T., 2011, JCAP, 1107, 034
Bouwens R. J., et al., 2014, Astrophys. J., 793, 115
Bouwens R., et al., 2015, Astrophys. J., 803, 34
Bouwens R. J., Oesch P. A., Stefanon M., Illingworth G., Labbé I., et al., 2021, Astrophys. J., 162, 47
Bouwens R., Illingworth G., Oesch P., Stefanon M., Naidu R., van Leeuwen I., Magee D., 2023, MNRAS, 523, 1009
Bromm V., Coppi P. S., Larson R. B., 2002, ApJ, 564, 23
Casey C. M., et al., 2022
Castellano M., Fontana A., Treu T., Santini P., Merlin E., et al., 2022, Astrophys. J. Lett., 938, L15
Ceverino D., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 2791
Cooray A., Sheth R. K., 2002, Phys. Rept., 372, 1
Corasaniti P., Agarwal S., Marsh D., Das S., 2017, Phys. Rev. D, 95, 083512
Curtis-Lake E., et al., 2022, arXiv e-prints, p. arXiv:2212.04568
Dekel A., Sarkar K. C., Birnboim Y., Mandelker N., Li Z., 2023, MNRAS, 523, 3201
Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rept., 733, 1
Donnan C. T., McLeod D. J., McLure R. J., Dunlop J. S., Carnall A. C., et al., 2023, Mon. Notices Royal Astron. Soc., 520, 4554
Eisenstein D. J., et al., 2023, arXiv e-prints, p. arXiv:2306.02465
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, Mon. Not. Roy. Astron. Soc., 406, 2267
Faucher-Giguère C.-A., 2018, MNRAS, 473, 3717
Ferrara A., Pallottini A., Dayal P., 2022, arXiv e-prints, p. arXiv:2208.00720
Finkelstein S. L., 2016, Publ. Astron. Soc. Australia, 33, e037
Finkelstein S. L., Bagley M. B., 2022, ApJ, 938, 25
Finkelstein S. L., et al., 2022a, arXiv e-prints, p. arXiv:2211.05792
Finkelstein S. L., et al., 2022b, ApJ, 940, L55
Foreman-Mackey D., 2016, The Journal of Open Source Software, 1, 24
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publ. Astron. Soc. Pac., 125, 306
Fujimoto S., et al., 2023, ApJ, 949, L25
Furlanetto S. R., Mirocha J., 2022a, arXiv e-prints, p. arXiv:2208.12828
Furlanetto S. R., Mirocha J., 2022b, MNRAS, 511, 3895
Furlanetto S. R., Mirocha J., Mebane R. H., Sun G., 2017, MNRAS, 472, 1576
Giavalisco M., Dickinson M., 2001, Astrophys. J., 550, 177
Hadzhiyska B., Liu S., Somerville R. S., Gabrielpillai A., Bose S.,

Eisenstein D., Hernquist L., 2021, Mon. Not. Roy. Astron. Soc., 508, 698
Harikane Y., et al., 2018, PASJ, 70, S11
Harikane Y., Ono Y., Ouchi M., Liu C., Sawicki M., et al., 2022, Astrophys. J., Suppl. Ser., 259, 20
Harikane Y., Nakajima K., Ouchi M., Umeda H., Isobe Y., Ono Y., Xu Y., Zhang Y., 2023a, arXiv e-prints, p. arXiv:2304.06658
Harikane Y., Ouchi M., Oguri M., Ono Y., Nakajima K., et al., 2023b, Astrophys. J., Suppl. Ser., 265, 5
Inayoshi K., Harikane Y., Inoue A. K., Li W., Ho L. C., 2022, ApJ, 938, L10
Jungman G., Kamionkowski M., Kosowsky A., Spergel D. N., 1996, Phys. Rev. D, 54, 1332
Kauffmann O. B., et al., 2022, A&A, 667, A65
Kocevski D. D., et al., 2023, ApJ, 946, L14
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottloeber S., Allgood B., Primack J. R., 2004, Astrophys. J., 609, 35
La Plante P., Mirocha J., Gorce A., Lidz A., Parsons A., 2023, Astrophys. J., 944, 59
Lee K., Giavalisco M., Gnedin O. Y., Somerville R., Ferguson H., Dickinson M., Ouchi M., 2006, Astrophys. J., 642, 63
Leung G. C. K., et al., 2023, arXiv e-prints, p. arXiv:2306.06244
Livermore R., Finkelstein S., Lotz J., 2017, Astrophys. J., 835, 113
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Madau P., Dickinson M., 2014, Ann. Rev. Astron. Astrophys., 52, 415
Maiolino R., et al., 2023, arXiv e-prints, p. arXiv:2305.12492
Mason C. A., Naidu R. P., Tacchella S., Leja J., 2019, Mon. Not. Roy. Astron. Soc., 489, 2669
Mason C. A., Trenti M., Treu T., 2023, Mon. Not. Roy. Astron. Soc., 521, 497
Menci N., Grazian A., Lamastra A., Calura F., Castellano M., et al., 2018, Astrophys. J., 854, 1
Meurer G. R., Heckman T. M., Calzetti D., 1999, Astrophys. J., 521, 64
Mirocha J., 2020, MNRAS, 499, 4534
Mirocha J., Furlanetto S. R., 2023, Mon. Notices Royal Astron. Soc., 519, 843
Mirocha J., Furlanetto S. R., Sun G., 2017, MNRAS, 464, 1365
Mirocha J., La Plante P., Liu A., 2021, Mon. Not. Roy. Astron. Soc., 507, 3872
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, Mon. Not. Roy. Astron. Soc., 428, 3121
Muñoz J. B., 2023, MNRAS, 523, 2587
Naidu R. P., et al., 2022, ApJ, 940, L14
Neistein E., van den Bosch F. C., 2006, Mon. Not. Roy. Astron. Soc., 372, 933
Oesch P. A., Brammer G., van Dokkum P. G., Illingworth G. D., Bouwens R. J., et al., 2016, Astrophys. J., 819, 129
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Overzier R. A., Bouwens R. J., Illingworth G. D., Franx M., 2006, Astrophys. J. Lett., 648, L5
Padmanabhan H., Loeb A., 2023, arXiv e-prints, p. arXiv:2306.04684
Park J., Mesinger A., Greig B., Gillet N., 2019, Mon. Not. Roy. Astron. Soc., 484, 933
Pérez-González P. G., et al., 2023
Ren K., Trenti M., Mutch S. J., 2018, ApJ, 856, 81
Ren K., Trenti M., Mason C. A., 2019, The Astrophysical Journal, 878, 114
Robertson B. E., 2010, ApJ, 716, L229
Rodriguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, Mon. Not. Roy. Astron. Soc., 462, 893
Rudakovskyi A., Mesinger A., Savchenko D., Gillet N., 2021, MNRAS, 507, 3046
Sabti N., Muñoz J. B., Blas D., 2021, JCAP, 01, 010
Sabti N., Muñoz J. B., Blas D., 2022a, Phys. Rev. D, 105, 043518
Sabti N., Muñoz J. B., Blas D., 2022b, Astrophys. J. Lett., 928, L20
Sabti N., Muñoz J. B., et al., 2023a, In prep
Sabti N., Muñoz J. B., Kamionkowski M., 2023b, arXiv e-prints, p. arXiv:2305.07049
Schneider A., Giri S. K., Mirocha J., 2021, Phys. Rev. D, 103, 083025
Shen X., Vogelsberger M., Boylan-Kolchin M., Tacchella S., Kannan R., 2023, arXiv e-prints, p. arXiv:2305.05679
Sheth R. K., Tormen G., 2002, Mon. Not. Roy. Astron. Soc., 329, 61
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
Steinhardt C. L., Sneppen A., Mostafa B., Hensley H., Jermyn A. S., et al., 2022, Astrophys. J., 931, 58
Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, Astrophys. J., 868, 92
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlober S., 2010, Astrophys. J., 724, 878
Trac H., Cen R., Mansfield P., 2015, Astrophys. J., 813, 54
Trapp A. C., Furlanetto S. R., Yang J., 2022, Mon. Not. Roy. Astron. Soc., 510, 4844
Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, Astrophys. J. Lett., 714, L202
Treu T., et al., 2022, Astrophys. J., 935, 110
Wang Y., et al., 2022, Astrophys. J., 928, 1
Waters D., Di Matteo T., Feng Y., Wilkins S. M., Croft R. A. C., 2016, MNRAS, 463, 3520
Wechsler R. H., Tinker J. L., 2018, Ann. Rev. Astron. Astrophys., 56, 435
Xavier H. S., Abdalla F. B., Joachimi B., 2016, Mon. Not. Roy. Astron. Soc., 459, 3693
Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2018, Monthly Notices of the Royal Astronomical Society, 483, 2983–3006
Yung L. Y. A., et al., 2022, MNRAS, 515, 5416
Yung L. Y. A., Somerville R. S., Finkelstein S. L., Wilkins S. M., Gardner J. P., 2023, arXiv e-prints, p. arXiv:2304.04348
Zavala J. A., et al., 2023, ApJ, 943, L9
Zentner A. R., Hearin A., van den Bosch F. C., Lange J. U., Villarreal A. S., 2019, MNRAS, 485, 1196
Zheng Z., Weinberg D. H., 2007, Astrophys. J., 659, 1
van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science and Engineering, 13, 22

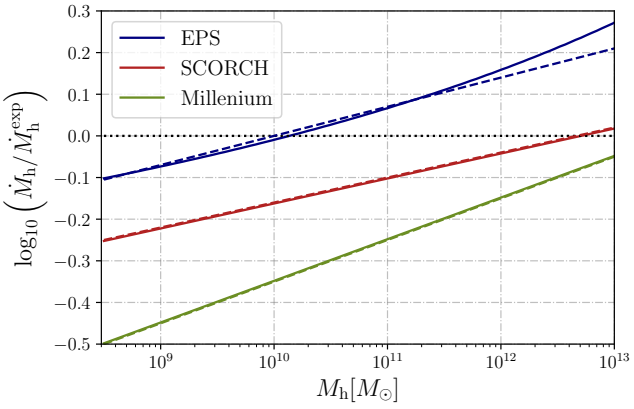

Figure A1. Accretion rates at $z = 10$ calibrated to different simulations, normalized by the exponential model assumed in the main text. Dashed curves show (log-)linear fits, which match very well, meaning that different $\dot{M}_{\rm h}$ fits just translate into a shift of the $f_\star$ parameters (and provide equal predictions).

APPENDIX A: ACCRETION RATES

Our model assumes that the star-formation rate is the product of the (gas) accretion rate $\dot{M}_{\rm g} = f_{\rm b}\dot{M}_{\rm h}$ and the star-formation efficiency $f_\star$. While there is no universal formula for the mass-accretion rate $\dot{M}_{\rm h}$ in the literature, here we argue that different functional forms are equivalent to each other, as their differences can be absorbed into the four parameters that determine $f_\star(M_{\rm h})$ (which are varied in our MCMCs).

Throughout the text, we assumed exponential accretion, where $\mathrm{d}M_{\rm h}/\mathrm{d}z = a_{\rm acc}M_{\rm h}$ and $a_{\rm acc} = 0.79$, is calibrated to simulations (Schneider et al. 2021). Other common prescriptions in the literature include the extended Press-Schechter formalism (Neistein & van den Bosch 2006), a fit to the `Millenium` simulations (Fakhouri et al. 2010), as well as a fit to the `SCORCH` simulations (Trac et al. 2015). We show these different accretion rates, divided by the exponential prescription, in Fig. A1. This figure is at $z = 10$, but is virtually identical at other z in the range of interest. For each curve, we present a linear approximation, which agrees remarkably well over the range of masses we consider. This means that the differences with respect to the exponential model can be reabsorbed into the mass dependence of $f_\star$ (in particular the intercept of the linear fit will renormalize $\epsilon_{\star,\rm UV}$, and the slope will shift $\alpha_\star$ and $\beta_\star$ simultaneously). As such, these different mass-accretion histories can be recast as a slightly different average SFE, otherwise providing identically good fits to all high-z data.

APPENDIX B: STOCHASTICITY THROUGH A DUTY CYCLE

In the main text, we have focused on how stochasticity can manifest through the variance of the halo-galaxy connection $P(M_{\rm UV}|M_{\rm h})$, given by:

$$P(M_{\rm UV}|M_{\rm h}) = \frac{1}{\sqrt{2\pi\sigma^2_{\rm UV}}}\exp\left[-(M_{\rm UV}-\overline{M_{\rm UV}})^2/(2\sigma^2_{\rm UV})\right] \ , \tag{B1}$$

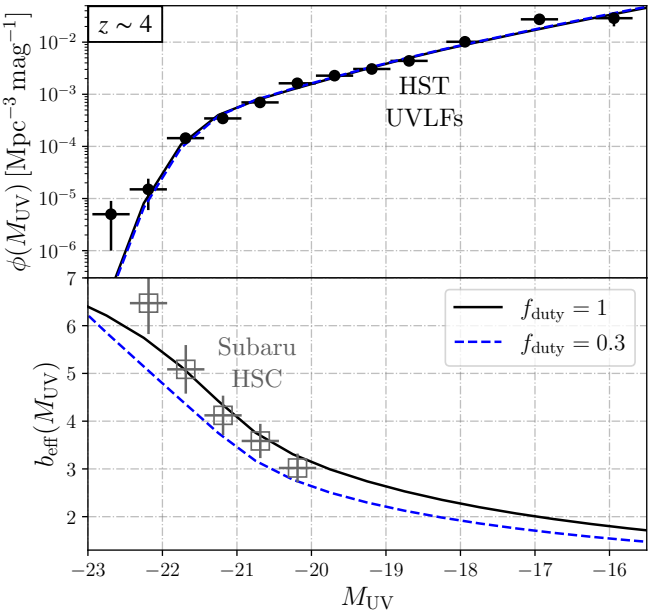

Figure B1. Same as Fig. 1, but for a model with a variable duty cycle $f_{\rm duty}$ (with fixed $\sigma_{\rm UV} = 0.3$). Lowering $f_{\rm duty}$ can be compensated by adjusting the SFE $f_\star$ to provide a good fit to the UVLF, which however results in different bias predictions.

with width $\sigma_{\rm UV}$ and mean $\overline{M_{\rm UV}}(M_{\rm h})$. A larger $\sigma_{\rm UV}$ "blurs" the halo-galaxy connection, allowing low-mass halos to host luminous objects and vice-versa[2].

We also consider an alternative parameterization, where only a fraction $f_{\rm duty} \leq 1$ of galaxies are UV-bright at any point. This can be achieved by reducing $P(M_{\rm UV}|M_{\rm h})$ by a factor of $f_{\rm duty}$, and physically will occur whenever the duty cycle of galaxies is not unity. We show in Fig. B1 the UVLFs and bias predicted for two models, one with $f_{\rm duty} = 1$ (as in the main text) and one with $f_{\rm duty} = 0.3$. In the latter case, we have adjusted the rest of parameters in the halo-galaxy connection by hand to recover agreement in the UVLFs. Despite the one-point functions being nearly identical, the biases predicted by these two models are significantly different. The model with $f_{\rm duty} = 0.3$ boasts a smaller bias, as each galaxy ought to be more luminous to account for the same UVLF (and thus will tend to reside in smaller mass halos). *Subaru* HSC measurements mildly prefer $f_{\rm duty} \approx 1$ at $z = 4$, though as in the main text, we warn the reader that the $b_{\rm eff}$ data from HSC cannot be directly compared against our predictions, as it was derived from an HOD model with a different cosmology. This shows that clustering measurements are promising not only to determine $\sigma_{\rm UV}$, but also the duty cycle of galaxies.

[2] We note, in passing, that the average luminosity of galaxies shifts with $\sigma_{\rm UV}$, as a Gaussian distribution in $M_{\rm UV}$ is log-normal in $L_{\rm UV}$, and thus will have a larger mean (Xavier et al. 2016). We have tested that the degeneracy remains after canceling this enhancement.

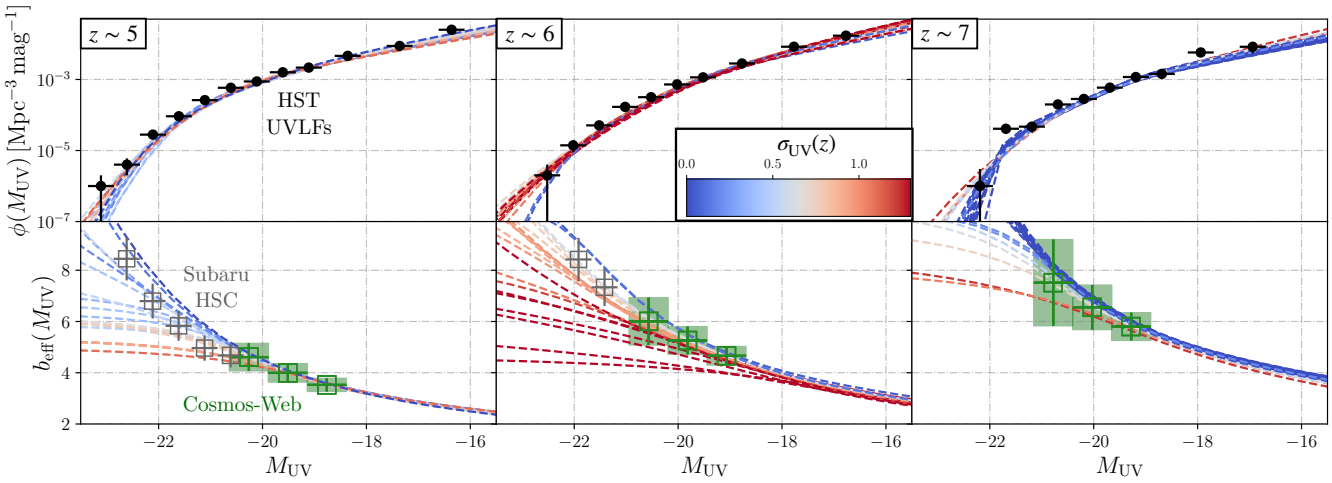

Figure D1. Same as Fig. 1, but for $z = 5-7$. We have fit the astrophysical model at each z independently, and show lines with parameters sampled from the 2σ preferred region at each z, colored by their value of $\sigma_{\rm UV}$. Along with the reported biases from *Subaru* HSC (gray squares), we show the forecast for a Cosmos-Web-like JWST survey in green, which reaches deeper magnitudes and higher z.

APPENDIX C: BIAS DATA FROM HSC

Here we describe how we derived the binned bias measurements from HSC that we show in this work. These have been derived from the HOD analysis in Harikane et al. (2022), which obtained $b_{\rm eff}(m_{\rm UV} < m_{\rm cut})$ for different UV magnitude cuts $m_{\rm cut}$. In order to translate these into binned (and number-weighted) biases, we take:

$$b_{\rm eff}(m_{\rm UV} \in [m_{\rm cut,1}, m_{\rm cut,2}]) = f_1 b_{\rm eff}(m_{\rm cut,1}) - f_2 b_{\rm eff}(m_{\rm cut,2}) \ , \quad \text{(C1)}$$

for $f_i = n_i/(n_1 - n_2)$, where n_i are the number of objects in each cut-off from Harikane et al. (2022). Given the lack of error bars in n_i, and the fact that these biases are derived with an HOD model within a different fiducial cosmology, we always show the biases with a 10% minimum error-bar. In future work, we will use the full angular correlation function information from the *Subaru* HSC (Sabti et al. 2023a).

APPENDIX D: CLUSTERING AT INTERMEDIATE REDSHIFTS

In this appendix, we extend the analysis from Fig. 1 to the $z = 5-7$ range, each fit independently to the data from Bouwens et al. (2021), see Fig. D1. Each z suffers from the same SFE-$\sigma_{\rm UV}$ degeneracy that we studied in the main text, so their 2σ confidence intervals cover a broad swath of values of $\sigma_{\rm UV}$. As we saw for the $z = 4$ case, more stochasticity (larger $\sigma_{\rm UV}$) results in lower biases that flatten towards the bright end. This is in conflict with the bias measurements from HSC (Harikane et al. 2022), which seem to disfavor $\sigma_{\rm UV} \gtrsim 1$ at $z \sim 5$ and 6. Moreover, we illustrate how a Cosmos-Web-like JWST survey can push the HSC results to fainter magnitudes and higher z. This will be key to unveil the nature of the halo-galaxy connection at high redshifts.

	HST ($\epsilon_{\star,\rm UV}$)	+JWST	HST ($\sigma_{\rm UV}$)	+JWST
$\alpha_\star$	0.61	0.84	0.74	0.69
$\mathrm{d}\alpha_\star/\mathrm{d}z$	−0.01	0.05	0.03	0.01
$\beta_\star$	−1.91	−1.26	−1.76	−1.68
$\mathrm{d}\beta_\star/\mathrm{d}z$	0.08	0.14	−0.02	0.18
$\log_{10} M_c$	12.03	11.83	11.84	11.93
$d\log_{10} M_c/dz$	0.03	−0.03	−0.02	0.00
$\log_{10}\epsilon$	–	–	−1.08	−1.11
$\mathrm{d}\log_{10}\epsilon/\mathrm{d}z$	–	–	−0.07	−0.08
$\sigma_{\rm UV}$	0.65	0.59	–	–
$\mathrm{d}\sigma_{\rm UV}/\mathrm{d}z$	−0.03	−0.03	–	–

Table E1. Best-fit values for the astrophysical parameters used in the main text, as defined in Eq. (4). Results are shown for two models, one in which we vary $\epsilon \equiv \epsilon_{\star,\rm UV}$ at each z (left two columns, reported in blue in Fig. 3) and one in which we do the same for $\sigma_{\rm UV}$ (right two columns, red in Fig. 3); for an analysis with HST only or HST + JWST data. The rest of parameters are assumed to vary linearly with z in each case, and we report their values and derivatives at $z = 8$.

APPENDIX E: BEST-FIT PARAMETERS

In this appendix, we provide the best-fit values used throughout the text for the astrophysical parameters and outline our MCMC approach for ease of reproducibility.

We have used the `emcee` code (Foreman-Mackey et al. 2013), and run chains with the likelihood and priors described in the main text, where the UVLFs are predicted with the public `Zeus21` code (Muñoz 2023). For each result, we have run 3.6×10^5 points (separated into 36 walkers), of which the first fifth are tossed out as burn-in. Each point takes ~ 1 ms to run, so it takes a chain a few hours to converge. We show our best-fit parameters for the four cases considered in the text in Tab. E1.