LUMINOUS RED GALAXIES: SELECTION AND CLASSIFICATION BY COMBINING OPTICAL AND INFRARED PHOTOMETRY

ABHISHEK PRAKASH, TIMOTHY C. LICQUIA, JEFFREY A. NEWMAN, AND SANDHYA M. RAO
PITT PACC, Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA; abp15@pitt.edu

ABSTRACT

We describe a new method of combining optical and infrared photometry to select luminous red galaxies (LRGs) at redshifts $z > 0.6$. We explore this technique using a combination of optical photometry from CFHTLS and *Hubble Space Telescope*, infrared photometry from the *Wide-Field Infrared Survey Explorer* satellite, and spectroscopic or photometric redshifts from the DEEP2 Galaxy Redshift Survey or COSMOS. We present a variety of methods for testing the success of our selection and present methods for optimization given a set of rest-frame color and redshift requirements. We have tested this selection in two different regions of the sky, the COSMOS and Extended Groth Strip (EGS) fields, to reduce the effect of cosmic/sample variance. We have used these methods to assemble large samples of LRGs for two different ancillary programs as a part of the Sloan Digital Sky Survey (SDSS)-III/Baryon Oscillation Spectroscopic Survey. This technique is now being used to select ∼600,000 LRG targets for SDSS-IV/Extended Baryon Oscillation Spectroscopic Survey, which began observations in Fall 2014, and will be adapted for the proposed Dark Energy Spectroscopic Instrument survey. We have found that these methods can select high-redshift LRGs efficiently with minimal stellar contamination; this is extremely difficult to achieve with selections that rely on optical photometry alone.

Key words: catalogs – cosmology: observations – galaxies: distances and redshifts – galaxies: general – galaxies: photometry – methods: data analysis

1. INTRODUCTION

Luminous red galaxies (LRGs) are relatively massive ($\sim 10^{11-12}\ M_\odot$), generally elliptical systems composed primarily of old stars; they are the most massive and luminous ($\geqslant 3\ L^\star$) galaxies in the $z \lesssim 1$ universe. These galaxies are expected to reside in massive dark matter halos and thus cluster very strongly (e.g., Padmanabhan et al. 2007). Their strong clustering enhances the baryon acoustic oscillation (BAO) signal. Combined with their intrinsic brightness, this makes them excellent probes of the large-scale-structure of the universe and a vital tool for cosmology.

LRGs exhibit a strong 4000 Å break in their spectral energy distributions (SEDs; Eisenstein et al. 2005). At lower redshifts ($z \lesssim 0.6$), LRGs can be efficiently selected and their redshifts estimated using optical photometry alone, by taking advantage of this feature. This method was used to select LRG targets for the Sloan Digital Sky Survey (SDSS) and the SDSS-III/Baryon Oscillation Spectroscopic Survey (BOSS), as well as the 2dF-SDSS LRG and QSO survey (2SLAQ; Eisenstein et al. 2001; Cannon et al. 2006). However, this method becomes extremely difficult at greater distances as cosmic expansion redshifts the 4000 Å break into longer wavelength filters, therefore requiring long imaging exposure times to overcome the brightness of the night sky in the near-infrared (NIR). A new method is required in order to efficiently select LRGs at higher redshifts.

Old stellar populations exhibit global maxima in their SEDs at a rest-frame wavelength of 1.6 μm, corresponding to the minimum in the opacity of H ions in their stellar atmospheres (John 1988). Since the light measured from LRGs is predominantly produced by old stars, we expect this feature to dominate their overall SEDs. This enables efficient selection of LRGs at higher redshifts.

In this paper, we demonstrate that a simple cut in optical–infrared color–color space provides an efficient method for differentiating LRGs from other types of objects. The methods described here will be applied for selecting LRG targets for next-generation spectroscopic surveys like the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) and the Dark Energy Spectroscopic Instrument (DESI) Survey and may be easily adapted to meet the needs of future prospects.

This paper is organized as follows: in Section 2, we describe the construction of samples used to test LRG selection methods, including both the imaging and spectroscopic data sets used. In Section 3, we present a simple selection method for identifying LRGs based on optical and infrared photometry and analyze the efficiency of this selection method for a set of nominal selection cuts applied to our sample. In Section 4, we explore methods for optimizing the LRG selection algorithm by adjusting the parameters of our cuts in color–color space. In Section 5, we summarize our results and conclude with plans for future work. For this work, we assume a standard ΛCDM cosmology with $H_0 = 100\ h\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2. DATA

In this paper, we make use of five cross-matched catalogs that cover two different regions of the sky, the Extended Groth Strip (EGS), with $214^\circ.0 < \alpha < 215^\circ.70$ and $52^\circ.14 < \delta < 53^\circ.22$, and the COSMOS field, with $149^\circ.41 < \alpha < 150^\circ.82$ and $1^\circ.49 < \delta < 2^\circ.92$. These two regions have been surveyed by a variety of telescopes, providing photometry over a wide range of wavelengths. We have cross-matched objects based on their positions on the sky as recorded by each survey. In order to avoid duplicate matches, we match each object in the catalog with the lowest surface density to its nearest neighbors in the denser catalogs that are closer than 1″.5. For DEEP2 objects in the EGS, we use the cross-identifications provided by

Matthews et al. (2013). Below, we briefly describe each of the catalogs used in this study.

2.1. Optical Photometry

Canada–France–Hawaii Telescope Legacy Survey (CFHT LS): CFHT LS consists of two parts. The Wide Survey covered $\sim$150 deg^2 divided over four fields with magnitude limits (50% completeness for point sources) of $u^\star \sim 26.0$, $g' \sim 26.5$, $r' \sim 25.9$, $i' \sim 25.7$, and $z' \sim 24.6$. The Deep Survey consists of four fields of 1 deg^2 area each, each with magnitude limits of $u^\star \sim 27.5$, $g' \sim 27.9$, $r' \sim 27.7$, $i' \sim 27.4$, and $z' \sim 26.2$. We use both the D2 Deep field, which lies within the COSMOS region, and D3, which overlaps with EGS. We also use both the Wide survey and the Deep survey in the EGS.

We use the COSMOS $ugriz$ magnitudes and their corresponding errors from the CFHT LS catalogs produced by Gwyn (2012), which were created using the *MegaPipe* data pipeline at the Canadian Astronomy Data Centre.

CFHT LS uses MegaCam filters that are slightly redder than their SDSS counterparts. We convert CFHT LS photometry to SDSS passbands by inverting the filter relations given in Equations (1)–(5). The relations for the g, r, i, and z bands come from analyses by the SuperNova Legacy Survey group.[1] The relation for the u band is taken from the CFHT web pages.[2] The transformed equations are

$$u_{\mathrm{SDSS}} = u_{\mathrm{Mega}} + 0.181\left(u_{\mathrm{Mega}} - g_{\mathrm{Mega}}\right), \quad (1)$$

$$g_{\mathrm{SDSS}} = g_{\mathrm{Mega}} + 0.195\left(g_{\mathrm{Mega}} - r_{\mathrm{Mega}}\right), \quad (2)$$

$$r_{\mathrm{SDSS}} = r_{\mathrm{Mega}} + 0.011\left(g_{\mathrm{Mega}} - r_{\mathrm{Mega}}\right), \quad (3)$$

$$i_{\mathrm{SDSS}} = i_{\mathrm{Mega}} + 0.001\left(r_{\mathrm{Mega}} - i_{\mathrm{Mega}}\right), \quad (4)$$

$$z_{\mathrm{SDSS}} = z_{\mathrm{Mega}} + 0.099\left(i_{\mathrm{Mega}} - z_{\mathrm{Mega}}\right), \quad (5)$$

where u_{Mega}, g_{Mega}, r_{Mega}, i_{Mega}, and z_{Mega} represent the $ugriz$ magnitudes measured by CFHT LS and u_{SDSS}, g_{SDSS}, r_{SDSS}, i_{SDSS}, and z_{SDSS} are the standard SDSS magnitudes. The resulting SDSS-passband magnitudes are then corrected for Galactic extinction using the dust map of Schlegel et al. (1998, hereafter SFD). To calculate the extinction in a given band, $A(\lambda)$, we interpolate the standard total-to-selective extinction ratios, i.e., $A(\lambda)/E(B-V)$ from Table 6 of Schlegel et al. (1998) for the effective wavelengths given in the filter list of CFHT LS.[3] We obtain $E(B-V)$ values from the SFD dust map (Schlegel et al. 1998) via the routine *dust_getval.pro* provided in the idlutils package.[4]

2.2. Infrared Photometry

Wide-Field Infrared Survey Explorer (*WISE*) catalog: *WISE* completed a mid-infrared survey of the entire sky by 2010 July in four infrared channels, labeled *W1, W2, W3*, and *W4*, centered at 3.4, 4.6, 12, and 22 μm, respectively. This was achieved using a 40 cm telescope with much higher sensitivity

than previous infrared survey missions. *WISE* achieved 5σ point-source sensitivities better than 0.08, 0.11, 1, and 6 mJy, corresponding to 19.1423, 18.7966, 16.4001, and 14.4547 *AB* magnitudes,[5] with angular resolutions of 6$\farcs$1, 6$\farcs$4, 6$\farcs$5, and 12$\farcs$0 in the *W1*, *W2*, *W3*, and *W4* channels, respectively (Wright et al. 2010).

A detection by *WISE* is required for an object to be in our catalog. This restriction will have negligible effect on LRGs, since they are bright in the *W1* band, but greatly reduces the number of objects to which we must apply our selection cuts. Based on the color–magnitude diagram, we observe that all $z < 20.5$ LRGs are detected in *W1* band at greater than 5σ. The 3.4 μm (*W1*) magnitudes are taken from the publicly available *WISE* All-Sky Data Release catalog of Wright et al. (2010). We convert these to the *AB* magnitude system and correct them for reddening using the SFD dust map (Schlegel et al. 1998) and interpolate extinction ratios, much as above.

2.3. Redshifts

COSMOS: the COSMOS photometric redshift ("photo-z") catalog from Ilbert et al. (2009) is a magnitude-limited catalog with $I < 25$. This catalog provides photometric redshifts over the $\sim$2 deg^2 COSMOS field. The redshifts are computed using 30 bands covering the UV (*GALEX*), visible–NIR (Subaru, CFHT, UKIRT), and mid-IR (*Spitzer*/IRAC). A χ^2 template-fitting method yields photo-z estimates that are calibrated with spectroscopic redshift measurements from Very Large Telescope-VIMOS and Keck-DEIMOS. For details of photo-z determinations and accuracy, see Mobasher et al. (2007).

EGS: in the EGS we use the DEEP2 spectroscopic catalog. DEEP2 is a high-resolution redshift survey of $\sim$53,000 galaxies at redshifts $z \gtrsim 0.7$ using the DEIMOS spectrograph at Keck Observatory (Newman et al. 2013). The survey covers an area of 2.8 deg^2 over four different fields. DEEP2 targeted galaxies brighter than $R_{AB} \sim 24.1$ with a spectral resolution of $R (= \Delta\lambda/\lambda) \sim 6000$ and a central wavelength of 7800 Å (Newman et al. 2013).

Since the DEEP2 catalogs only provide *BRI* photometry, Matthews et al. (2013) have created a catalog to supplement them with *ugriz* photometry from CFHTLS and SDSS. Each catalog is cross-matched by position on the sky in order to assign *ugriz* photometry to objects in the DEEP2 catalogs. We use the Matthews et al. (2013) catalog in the EGS field to obtain *ugriz* photometry. We correct this photometry for extinction as described above in Section 2.1.

All objects in our data sets are required to have reliable redshifts. This is important as this information is used in determining the rest-frame colors of galaxies. For the COSMOS field, we use the photometric redshifts, zp_gal, taken directly from Ilbert et al. (2009). We do not consider objects with zp_best = NULL as these are the objects in the masked area. For the galaxies in the EGS, we make use of the heliocentric reference-frame spectroscopic redshift, *ZHELIO*, provided in the DEEP2 extended photometry catalog of Matthews et al. (2013). We also ensure that each galaxy in our sample has a securely measured redshift (i.e., we require redshift quality flags, *ZQUALITY* of 3 or 4 in DEEP2).

2.4. Object Type Identification

COSMOS: in order to distinguish stars, galaxies, and X-ray sources within the COSMOS field from each other, we apply a variety of cuts based on the photometric redshift estimates, *zp_best*, as well as the reduced chi-squared value associated with the separate star and galaxy template fits to each object's spectral energy distribution (SED), *Chi_star* and *Chi_gal*.[6] The parameters we use have been provided by Ilbert et al. (2009). To identify different objects, we use the criteria

$$\text{X-ray sources: } zp\_best > 9, \qquad (6)$$

$$\begin{aligned}\text{Galaxies: } & (Chi\_gal < Chi\_star) \\ \text{AND } & (0.011 < zp\_best < 9), \end{aligned} \qquad (7)$$

$$\begin{aligned}\text{Stars: } & (chi\_gal > chi\_star) \\ \text{OR } & (chi\_gal < chi\_star) \\ \text{AND } & (zp\_best < 0.011 \ OR \ zp\_best > 9). \end{aligned} \qquad (8)$$

These criteria distinguish galaxies from stars and X-ray sources. An object is flagged as a star if its SED is best fit with a stellar template or if it yields an extremely low redshift in the case where a galaxy template is the better fit. An object identified as a galaxy should not only fit the galaxy template best but also yield a redshift between 0.011 and 9. Once the objects are identified, *zp_final* is our redshift indicator and its value is set to the photometric redshift, *zp_gal*, if the object is identified as a galaxy. The *zp_final* value is set to 0.0 for stars and −9.99 for X-ray sources.

EGS: We use the *Hubble Space Telescope*–Advanced Camera for Surveys (*HST*–ACS) general catalog for objects in the EGS field. The *HST*–ACS General Catalog is a photometric and morphological catalog created using publicly available data obtained with the ACS instrument on the *HST*. This provides a large sample of objects with reliable structure measurements. It includes approximately 470,000 sources originally observed in a variety of sky surveys, including the All-Wavelength Extended Groth Strip International Survey, COSMOS, GEMS, and GOODS (Griffith et al. 2012). A single Sérsic model for each object is assumed for deriving quantitative structural parameters (e.g., surface brightness and effective radius).

Our goal is to be able to estimate the stellar contamination in the EGS. DEEP2 avoided targeting stars, so we cannot assess this from the spectroscopic sample on its own. No such effort is required for the COSMOS field since the catalog of Ilbert et al. (2009) contains both stars and galaxies. We use the same definition as Griffith et al. (2012) for identifying compact objects (presumed to be stars) based on their larger surface brightness, *mu_HI*, and lower effective half-light radius, *RE_Galfit_HI* (for more details, see Figure 5 of Griffith et al. 2012). Specifically, we identify objects with *mu_HI* < 18.5 or (*mu_HI* > 18.5 and *RE_Galfit_HI* < 0.03) as stars, where *RE_Galfit_HI* is given in arcseconds. Once the number of the stars is determined, we assume that the fraction of objects that are stars is uniform over the entire EGS. This gives us the total number of selected objects by our color cut (galaxies, X-ray sources, and stars), which is then used to calculate the normalized Figure of Merit (FOM); see Figure 12.

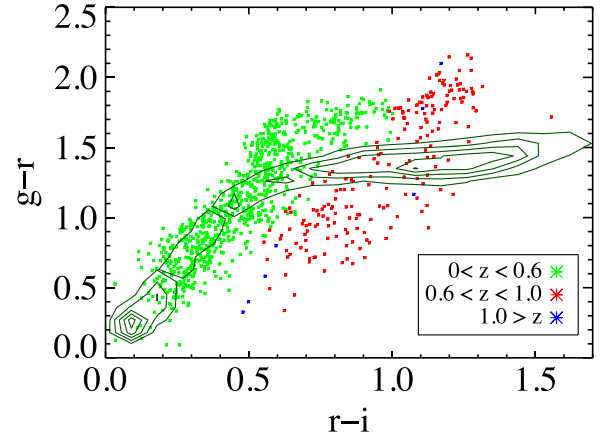6 Variables defined the same way they appear in the catalogs.

Figure 1. $g - r$ vs. $r - i$ optical color–color plot of galaxies (data points) observed by CFHT LS with COSMOS photometric redshifts. Star density contours are overplotted to show the overlap with galaxies. This overlap makes LRG selection difficult at higher redshifts, requiring a new method of selecting them.
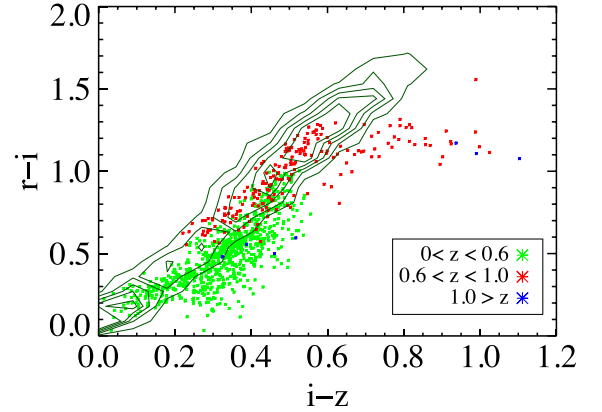

Figure 2. $r - i$ vs. $i - z$ optical color–color plot of galaxies (data points) observed by CFHT LS and the COSMOS survey, similar to Figure 2. Star density contours are overplotted to show the overlap with galaxies. Stars overlap almost entirely with high-redshift LRGs.

This step is necessary since the *HST*–ACS general catalog of Griffith et al. (2012) covers only a portion of the entire EGS.

3. METHOD OF LRG SELECTION

Our goal is to develop a method for selecting LRGs at high redshift, i.e., $z > 0.6$. One of the main challenges in LRG selection based on optical photometry alone is stellar contamination. The color overlap of stars with galaxies is illustrated in Figures 1 and 2, where contours depicting the density of stars are overlaid on the locations of galaxies (shown as dots) in $g - r$ versus $r - i$ (Figure 1) and $r - i$ versus $i - z$ (Figure 2) color–color plots. The strong overlap of these populations makes them difficult to separate cleanly.

In this paper, we present a new technique for identifying high-redshift LRGs that combines optical and infrared photometry. The lowest wavelength channel of imaging from the *WISE* satellite is centered around 3.4 μm. This overlaps with the redshifted "1.6 μm bump" at redshifts of $z \sim 0.6-1$, causing LRGs at those redshifts to appear very bright in this band. This phenomenon is illustrated in Figure 3, where a typical LRG at $z \sim 1.0$, which is barely detected in a $1' \times 1'$
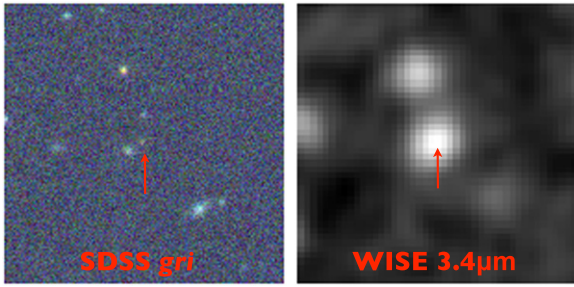
Figure 3. 1 arcmin square SDSS and *WISE* (3.4 μm) images of a $z \sim 1$ LRG. This object is at the 5σ detection limit in SDSS r and i but dominates the *WISE* image. Due to the redshifted "1.6 μm bump," LRGs at $z \sim 1$ look much brighter in the *WISE* 3.4 μm band than at optical wavelengths, providing a new method of selecting them.
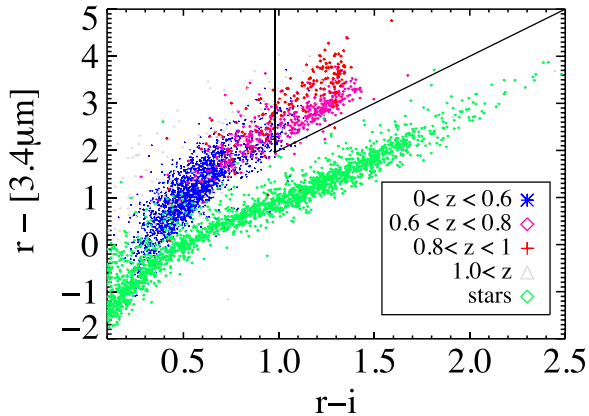

Figure 4. Optical/Infrared color–color plot for galaxies observed by *WISE*, CFHT LS, and the COSMOS survey. Blue symbols represent all the galaxies at $z < 0.6$. Pink diamonds represent galaxies at $z > 0.6$, while red crosses represent $z > 0.8$. Cyan triangles represent galaxies at higher redshift of $z > 1.0$. Stars are represented by green diamonds. The triangular region represents the LRG selection region used in Figures 6–8.

Figure 5. Standard color–magnitude diagram of DEEP2 DR4 galaxies in redshift range $0.5 < z < 1.0$. Rest-frame $U - B$ color is plotted on the y-axis, and the absolute B-band magnitude is plotted on the x-axis. We clearly see two different overdensities, one around $U - B \sim 0.4$–0.7 and another around $U - B \sim 1.25$. The galaxies with $U - B \geqslant 1.0$ are commonly referred to as the red sequence galaxies, while galaxies having $U - B < 1.0$ form the blue cloud. The low-density region in between these overdensities is generally referred to as the green valley. Rest-frame $U - B$ colors and absolute magnitudes are obtained using the k-correct package, which assumes $h = 1$; see Blanton & Roweis (2007) for details. Choosing a different value of h will not change $U - B$, just M_B.

region of SDSS optical imaging, is the brightest object in the *WISE* NIR image of the same area. The relative brightness of LRGs in the *WISE* 3.4 μm band compared to the optical bands increases monotonically up to $z \sim 1$ and then declines past $z \sim 1.1$ (at which point optically bright LRGs become rare).

To match the expected spectroscopic depth of DESI LRGs, we restrict the data set for all analyses in this paper to those objects that have SDSS z-band magnitude $z < 20.5$. We now present a new technique that combines both optical and infrared photometry as a means of selecting galaxies that are intrinsically red and at high redshift while circumventing most stellar contamination. In Figure 4, we show both stars and galaxies in a plot of $r - W1$ color based on *WISE* and SDSS-passband photometry as a function of their SDSS $r - i$ color. Here we can easily see that the two populations (stars, shown as green diamonds, and galaxies, shown as all other colored diamonds) exhibit a natural separation in the NIR–optical color space. In fact, the separation between the two populations grows as a function of galaxy redshift, allowing clean identification of the LRGs at higher redshifts, $z > 0.6$. Simultaneously, $r - i$ color increases with increasing redshift as the 4000 Å break shifts redward, particularly for intrinsically red galaxies, allowing a selection specifically for intrinsically red, higher-redshift objects.
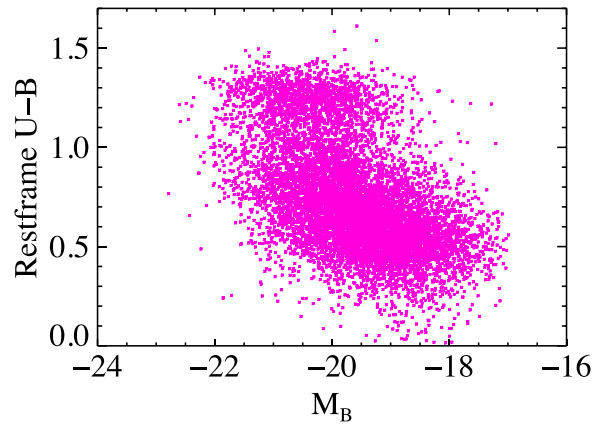
As a result, a simple cut in the optical–infrared color–color plot enables us to efficiently select LRGs at higher redshifts, rejecting bluer galaxies, lower-redshift objects, and stars. As a nominal scenario, we select all objects that have both $r - i > 0.98$ and $r - W1 > 2.0 \times (r - i)$, where r and i are extinction-corrected SDSS magnitudes and *W1* is the magnitude in the *WISE* 3.4 μm passband on the *AB* system (Figure 4). We have determined these cuts through visual optimization by examining the populations in Figure 4.

Overall, our LRG color-cut selection has three free parameters: the minimum allowed $r - i$ color (corresponding to the vertical line in Figure 4), and the slope and intercept of the line determining the minimum allowed $r - W1$ color at a given $r - i$ color (corresponding to the inclined line in Figure 4). The latter of these two criteria determines the degree to which stars are rejected from our sample of LRGs. The $r - i$ cut will mostly affect the properties of the galaxies we select (e.g., their redshift distribution). We investigate the performance of this color cut in Figures 6–8. For an object to be classified as a high-redshift LRG, we generally require it to have both a rest-frame color $U - B \geqslant 1$ and a redshift $z > 0.6$, though we also consider other redshift thresholds. We use the k-correct package to obtain rest-frame $U - B$ color for all galaxies; see Blanton & Roweis (2007) for details.

To further justify our choice of redness threshold, $U - B \geqslant 1$, we have plotted the standard color–magnitude diagram for DEEP2 galaxies in Figure 5, clearly showing the red sequence, blue sequence, and green valley. From Figure 6, we observe that with our nominal color cut, 85.8% of the galaxies selected have rest-frame $U - B \geqslant 1$, indicating that they are intrinsically red in rest-frame color, adopting the same criterion employed by Gerke et al. (2007). The remaining galaxies selected are still relatively red and massive but have some ongoing star formation.

In Figure 7, we show a histogram of the redshifts measured for each of the objects selected, indicating that 77.6% of the
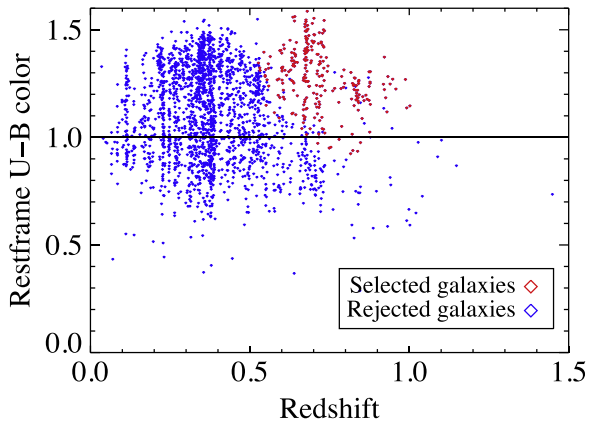
Figure 6. Rest-frame $U - B$ colors of *WISE*-detected, $z < 20.5$ galaxies as a function of redshift. Red dots represent the galaxies from the LRG selection region of Figure 4. Blue dots represent the galaxies excluded by this color selection. The red sequence corresponds well to those galaxies with $U - B \geqslant 1.0$. Of the galaxies selected by our color cut, 85.8% have $U - B$ greater than 1.0; most selected objects that miss this cut are only slightly bluer than the red sequence.
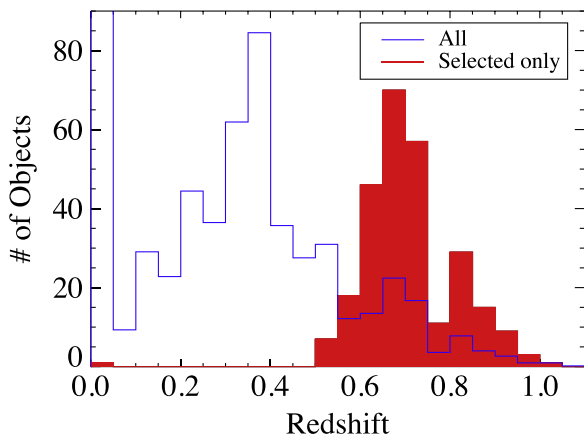

Figure 7. Redshift histogram of *WISE*-detected, $z < 20.5$ galaxies. The blue histogram depicts the redshift distribution of the full *WISE*-detected, $z < 20.5$ sample, with an arbitrary renormalization applied. The red histogram represents galaxies from the LRG selection region of Figure 4. The selected objects predominantly fall in the redshift range $0.6 < z < 1$, with a modest 0.63% contamination by stars (appearing at $z \sim 0$).

galaxies selected by our nominal cuts fall within the redshift range of interest, i.e., $0.6 \geqslant z \geqslant 1.0$, and <1% are stars.

4. OPTIMIZATION OF LRG SELECTION

4.1. Optimization Using Receiver Operating Characteristic

As a first attempt to optimize the efficiency of our selection method, we begin by varying the minimum allowed $r - i$ color (corresponding to shifting the vertical line in Figure 4), while keeping all other parameters fixed. We interpret the results of varying this color criterion using a receiver operating characteristic (ROC) plot, as shown in Figure 8. The ROC plot provides a visualization of the performance of a classification system; in our case, this quantifies our ability to segregate out the high-redshift LRGs in contrast to galaxies that are either blue or at lower redshift ("non-LRGs" for short). Each individual curve shows the result of using a different threshold on the minimum-allowed z to be a "high-redshift"

LRG; we only consider LRGs above the desired minimum redshift as our target population.

The y-axis of the ROC plot represents the true positive rate (TPR), also known as the "sensitivity." The TPR is defined as the fraction of all true high-redshift LRGs in the underlying sample that are within a given color cut. One of our main goals is to maximize the TPR. Of course, if the minimum-allowed $r - i$ color was shifted so blue as to select all galaxies, this would be achieved by definition. However, there is a cost associated with doing this: we would select many galaxies that are not LRGs. This misidentification is quantified as the false positive rate (FPR), which is plotted on the x-axis of the ROC plot. The FPR is the fraction of all non-LRGs (in our case, all $z < 20.5$, *WISE*-selected objects that are blue $[U - B < 1]$ and/or below the desired redshift) that are placed in the LRG sample by a given color cut. Our goal is to minimize this quantity when varying the color cuts used to pick our sample while at the same time maximizing the TPR.

One common way of assessing the performance of a selection algorithm is the area under curve (AUC) diagnostic (Hanley & McNeil 1982). This is calculated by integrating the ROC curve over all FPRs. Here, we assess the efficiency of selecting LRGs using three different redshift thresholds, $z \geqslant$ 0.55, 0.6, or 0.65. We have varied the minimum allowed $r - i$ color over the range in values from 0.8 to 1.2 and calculated the TPR and the FPR for each selection. Figure 8 shows that, with our chosen cuts, we are able to attain a TPR of 85%−95% (depending on the choice of the minimum allowed "high" redshift, threshold z) while at the same time keeping the FPR below ∼3%. Based on the AUC, we conclude that our selection algorithm performs best for a threshold redshift of $z \geqslant 0.6$ (corresponding to the blue curve in Figure 8), as it encompasses the maximum area.

4.2. Optimization Using FOM for Large-scale Structure Studies

To optimize the efficiency of our methods further, we attempt to account for the contribution to cosmological analyses from those non-LRG objects that are selected. While LRGs are the prime targets we are after, those galaxies that are not red and have a redshift of $z \geqslant 0.6$ (hereafter, "high-z blue galaxies") still provide useful information. To assess this, we define an FOM as

$$\mathrm{FOM} = a \times n_{\mathrm{LRGs}} + b \times n_{\mathrm{high}-z\ \mathrm{blue\ gals}}, \qquad (9)$$

where a and b are constants weighting the targets, and n_{LRGs} and $n_{\mathrm{high}-z\ \mathrm{blue\ gals}}$ represent the number density (number per unit area) of LRGs and high-redshift blue galaxies, respectively, for a given set of color cut criteria. Since LRGs are our prime targets, we assign $a = 1.0$ and $b < a$. For the purpose of this paper, we adopt $b = 0.75$ as an example; a more ideal weighting would set b/a according to the relative contribution of each class of objects toward, e.g., the uncertainty in the BAO scale. We assume that stars, X-ray sources (which tend to be at higher z), and galaxies at $z \leqslant 0.6$ contribute nothing toward the FOM. We want to optimize not only the total FOM but also the FOM per object, which we will refer to as the normalized FOM. The total FOM represents the total constraining power of the whole sample and will increase even when we select lower-value high-z blue galaxies. On the other hand, the normalized
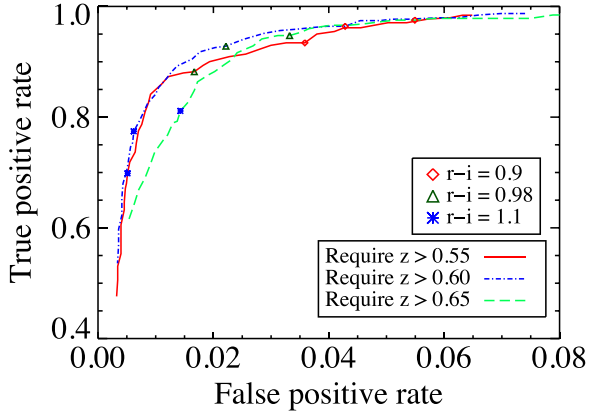
Figure 8. ROC plot for changes in the minimum allowed $r - i$ for selecting LRGs over the range $0.8 < r - i < 1.2$. For each curve, we only consider intrinsically red objects that have redshifts above a different specified minimum redshift, z, to be high-z LRGs. The green triangles represent the performance at the threshold of $r - i = 0.98$. Similarly, the red diamonds and the blue stars represent the performance at the threshold of $r - i = 0.9$ and $r - i = =1.1$, respectively. Based on the commonly used AUC criterion, our color cuts perform slightly better for LRGs with minimum redshift of $z \sim 0.6$ than 0.55 or 0.65.
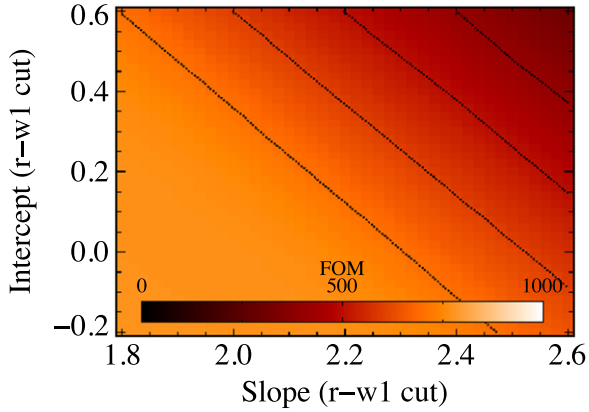
Figure 9. LRG selection FOM (see Equation (9)) as a function of the slope and intercept of the $r - W1$ cut, using a fixed $r - i$ threshold of 0.98. The visible pattern and the overplotted contours of constant FOM show that these parameters can be effectively replaced by a single linear combination of the two, as in Equation (10).

FOM will increase only when LRGs make up a higher fraction of the sample and, therefore, can be more useful for optimization. Both should be considered; selecting all the objects regardless of color would yield a large total FOM but little value per object, while an extremely restrictive selection could have normalized FOM ~ 1 but have little total constraining power due to the small number of objects chosen. We then create a three-dimensional grid to tabulate the FOM for each possible combination of the minimum allowed $r - i$ and the slope and intercept of the line marking the minimum allowed $r - W1$ for a given $r - i$ color. Figure 8 shows that our nominal cut of $r - i = 0.98$ performs quiet well on this metric.

To simplify the search space, we next fix our $r - i$ cut at the optimal value and analyze the FOM when varying the two parameters associated with the $r - W1$ cut. In Figure 9, we plot the FOM as a function of the slope and the intercept of the
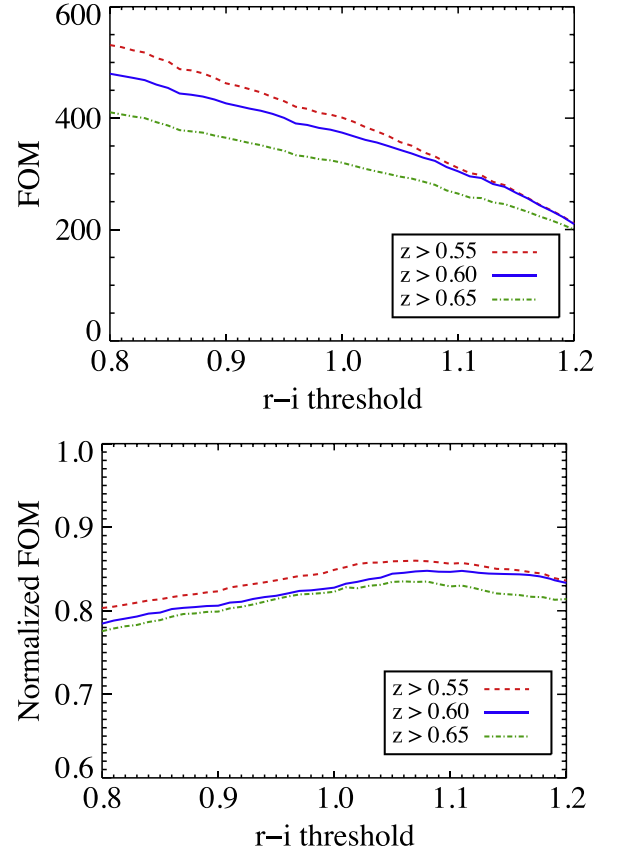

Figure 10. Maximum FOM for a given $r - i$ color threshold for $z < 20.5$ galaxies, using data in the COSMOS region. Each individual curve shows the result of using a different choice for the minimum-allowed z for a high-redshift LRG. In the top panel, the monotonic decrease in FOM as the threshold $r - i$ color moves redward (corresponding to moving the vertical line in Figure 4 to the right) is consistent with our expectations, since we select a decreasing number of both LRGs and high-z blue galaxies as the $r - i$ cut is moved to the right. In the bottom panel, we analyze the normalized FOM as the threshold $r - i$ color moves redward. For $r - i < 1.05$, as the color cut moves redder, it selects a higher fraction of high-z LRGs, increasing the purity of the sample and hence the normalized FOM monotonically.

$r - W1$ threshold line. We overplot the contours of constant FOM to highlight the linear nature. It is clear that the FOM depends primarily on a fixed combination of these parameters. As a result, the optimal value of the slope of the line, S_{r-W1}, given a value of the intercept, i_{r-W1}, can be obtained from the relation

$$S_{r-W1} = 2.0 - 0.4 \times i_{r-W1}. \qquad (10)$$

Although the exact form of this optimal relation is dependent on the value of b in Equation (9), the existence of this correlation is significant, as it reduces our selection algorithm from a three-parameter to a two-parameter problem. We find that similar linear correlations occur for different values of b in Equation (9) (e.g., $b = 0.25$ or 0.5). In contrast to FOM, the normalized FOM depends little on the slope/intercept in the relevant parameter range, so in this case, our decisions are driven by FOM.

Based on this approximation, we define a new variable v:

$$v = (2.0 - S_{r-W1}) - 0.4 \times i_{r-W1}. \qquad (11)$$

This is defined such that $v = 0$ at our nominal parameter values for the $r - W1$ cut. Next, we create a two-dimensional grid to
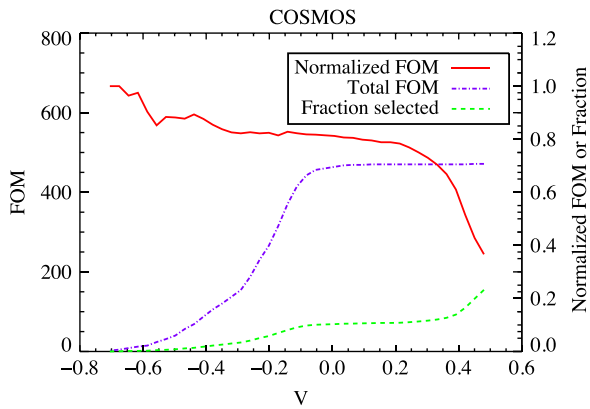
Figure 11. FOM at $r - i = 0.98$ for $z < 20.5$ galaxies as a function of the combined slope/intercept parameter (v) in the COSMOS region. Only galaxies above a threshold redshift of $z > 0.6$ are counted in the FOM. Each individual curve shows a different quantity. The second y-axis (on the right) is used for plotting both the normalized FOM (FOM normalized by the total number of objects selected by a given color cut) and the fraction of all $z < 20.5$ objects that are targeted. Values of v around 0 are roughly optimal for both the total and normalized FOM.

tabulate the FOM at each possible combination of the two parameters in our model, the minimum allowed $r - i$ and v. This grid is then analyzed to determine the parameter values that maximize the FOM. In Figure 10, we have plotted the maximum of the FOM over all i_{r-W1} values as a function of the $r - i$ color cut. The maximum FOM decreases monotonically as the threshold $r - i$ color moves redward (corresponding to moving the vertical line in Figure 4 to the right). This result is consistent with our expectation, since we select a decreasing number of both LRGs and high-z blue galaxies as the $r - i$ cut is moved redward.

In Figure 11, we have shown the FOM and normalized FOM as a function of v assuming $r - i = 0.98$. We have also overplotted the fraction of objects selected as a consistency check. The flatness of the curves in Figure 11 can be understood with the help of Figure 4, as well as Equations (10) and (11). As v becomes positive, the intercept of the $r - W1$ threshold line becomes negative. This region lies within the empty region separating stars and galaxies in Figure 4. Hence, we do not see any significant change in any of the quantities plotted. As v is further increased, the $r - W1$ cut starts including stars in our sample. This causes an abrupt increase in the fraction of objects selected and a corresponding abrupt decrease in the FOM normalized by the number of selected objects. However, the total FOM remains mostly unaffected as stars do not contribute to it.

Figure 8 illustrates that $r - i = 0.98$ is a well-optimized cut for our purposes. Any decrease in this parameter causes the FPR to increase significantly without any significant gain in the TPR. Overall, we conclude that our nominal cuts of $r - i > 0.98$ and $r - W1 > 2.0 \times (r - i)$ are well optimized and are the final cuts used in this selection. We find that when varying our threshold redshift marginally, e.g., $z > 0.55$ or $z > 0.65$, our analysis yields similar results for the optimal value of selection parameters.

The analyses described so far all rely on our COSMOS field sample. It is worthwhile to test whether adding more data from different regions of the sky, which will reduce sample/cosmic variance, can improve our optimization. As explained in
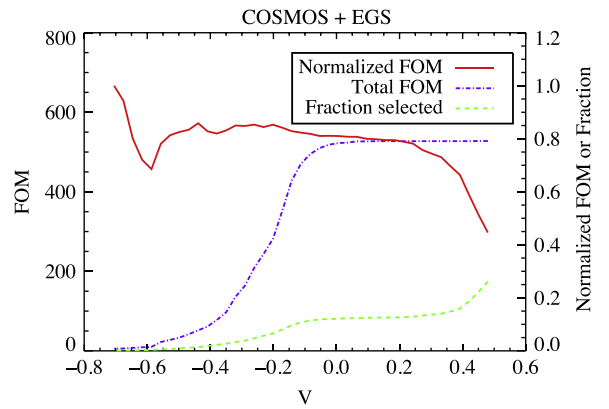

Figure 12. FOM for $r - i = 0.98$ as a function of the combined slope/intercept parameter (v), averaged over both the COSMOS and the EGS fields. Only galaxies above a threshold redshift of $z > 0.6$ are counted in the FOM. Similar to Figure 11, values of v around 0 prove to be roughly optimal for both the total and normalized FOM.

Section 2, in the EGS, we have obtained DEEP2 extended photometry from the catalog of Matthews et al. (2013). We repeat the same analysis as was done for the COSMOS field to estimate the FOM for our two-parameter model. However, to estimate the total number of objects selected by a given color cut, we need to estimate the stellar contamination. Since DEEP2 avoided targeting stars (Newman et al. 2013), this is done separately using the *HST*–ACS general catalog of Griffith et al. (2012), as described in Section 2.4. We otherwise repeat the same analysis as done for the COSMOS field to estimate the stellar contamination for a given color cut in 2D parameter space. The FOM in the EGS region shows a very similar behavior as in the COSMOS field, as can be seen in Figures 10 and 11, indicating that sample/cosmic variance is not a major issue.

In Figure 12, we show the behavior of the FOM averaged over both the EGS+COSMOS fields. The plot shows a very consistent behavior, enabling us to conclude that our baseline color selection, $r - i > 0.98$ and $r - W1 > 2.0 \times (r - i)$, is indeed well suited for selecting $z > 0.6$ LRGs using this method.

5. CONCLUSIONS AND FUTURE WORK

We have found a reliable and efficient method of identifying and selecting LRGs at higher redshifts by combining optical and infrared photometry. We have explored a variety of methods for optimizing our color cuts, given a particular set of rest-frame color and redshift requirements. With these optimization procedures, we can, for instance, tune the redshift range to select LRGs as required by different surveys.

These methods have now been used to assemble large samples of LRGs. More than 10,000 $z < 20$, SDSS+*WISE*-selected LRGs were targeted by a BOSS Ancillary program in 2012–2013 (SDSS DR12, 2015, in preparation). This will not only provide a good check on our selection methods but also greatly increase the sample that we can use to optimize the selection process further. We have also selected LRGs based on similar methods, but using colors derived only from SDSS i, z and *WISE W*1 (i.e., using $i - W1$ and $i - z$ colors). Selection in these redder bands helps for targeting higher-z LRGs, but they are not as efficient as the combination of r, i, and *W1* in star-galaxy separation.

We have created a sample of LRGs over the entire SDSS footprint that has been used in selecting targets for a second BOSS ancillary survey, the SDSS Extended Quasars, Emission line galaxy, and LRGs Survey (SEQUELS). In SEQUELS, we have targeted $\sim$70,000 $z < 20$ LRGs selected by one of two color cuts: one that utilizes $r - i$, $r - W1$, and a minimum value of $i - z$, and a second that uses only $i - z$ and $i - W1$ colors. The work of analyzing the resulting spectra is in progress and will be reported in future publications. The same methods are being used for selecting LRG targets for eBOSS, which began observations in Fall 2014. The results from these selection algorithms will be described a in future paper (A. Prakash et al. 2015, in preparation).

We are also investigating even deeper selections of LRGs using r, z, and $W1$ photometry for the proposed DESI survey (see DESI Conceptual Design Report).[7] Optical/IR LRG selections have proved to be effective in our tests and will provide a cornerstone sample for BAO surveys through the next decade.

REFERENCES

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734
Cannon, R., Drinkwater, M., Edge, A., et al. 2006, MNRAS, 372, 425
Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, AJ, 122, 2267
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Gerke, B. F., Newman, J. A., Faber, S. M., et al. 2007, MNRAS, 376, 1425
Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9
Gwyn, S. D. J. 2012, AJ, 143, 38
Hanley, J. A., & McNeil, B. J. 1982, Radiology, 143, 29
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
John, T. L. 1988, A&A, 193, 189
Matthews, D. J., Newman, J. A., Coil, A. L., Cooper, M. C., & Gwyn, S. D. J. 2013, ApJS, 204, 21
Mobasher, B., Capak, P., Scoville, N. Z., et al. 2007, ApJS, 172, 117
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Padmanabhan, N., Schlegel, D. J., Seljak, U., et al. 2007, MNRAS, 378, 852
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

---

[7] http://desi.lbl.gov/cdr/