DRAFT VERSION SEPTEMBER 16, 2008
Preprint typeset using LaTeX style emulateapj v. 11/27/05

SLAC-PUB-13730
July 2009

CONSTRAINING THE SCATTER IN THE MASS-RICHNESS RELATION OF MAXBCG CLUSTERS WITH WEAK LENSING AND X-RAY DATA

EDUARDO ROZO[1], ELI S. RYKOFF[2], AUGUST EVRARD[3,4,5], MATTHEW BECKER[6], TIMOTHY MCKAY[3,4,5], RISA H. WECHSLER[7], BENJAMIN P. KOESTER[8,9], JIANGANG HAO[3], SARAH HANSEN[8,9], ERIN SHELDON[10], DAVID JOHNSTON[11], JAMES ANNIS[12], JOSHUA FRIEMAN[8,9,12]

Draft version September 16, 2008

ABSTRACT

We measure the logarithmic scatter in mass at fixed richness for clusters in the maxBCG cluster catalog, an optically selected cluster sample drawn from SDSS imaging data. Our measurement is achieved by demanding consistency between available weak lensing and X-ray measurements of the maxBCG clusters, and the X-ray luminosity–mass relation inferred from the 400d X-ray cluster survey, a flux limited X-ray cluster survey. We find $\sigma_{\ln M|N_{200}} = 0.45^{+0.20}_{-0.18}$ (95% CL) at $N_{200} \approx 40$, where N_{200} is the number of red sequence galaxies in a cluster. As a byproduct of our analysis, we also obtain a constraint on the correlation coefficient between $\ln L_X$ and $\ln M$ at fixed richness, which is best expressed as a lower limit, $r_{L,M|N} \geq 0.85$ (95% CL). This is the first observational constraint placed on a correlation coefficient involving two different cluster mass tracers. We use our results to produce a state of the art estimate of the halo mass function at $z = 0.23$ — the median redshift of the maxBCG cluster sample — and find that it is consistent with the WMAP5 cosmology. Both the mass function data and its covariance matrix are presented.

Subject headings: galaxies: clusters – X-rays: galaxies: clusters - cosmology: observation

1. INTRODUCTION

The space density of galaxy clusters as a function of cluster mass is a well-known cosmological probe (see e.g. Holder et al. 2001; Haiman et al. 2001; Rozo et al. 2004; Lima & Hu 2004), and ranks among the best observational tools for constraining σ_8, the normalization of the matter power spectrum in the low redshift universe (see e.g. Frenk et al. 1990; Henry & Arnaud 1991; Schuecker et al. 2003; Gladders et al. 2007; Rozo et al. 2007b).[13] The basic idea is this: in the high mass limit, the cluster mass function falls off exponentially with mass, with the fall-off depending sensitively on the amplitude of the matter density fluctuations. Observing this exponential cutoff can thus place tight constraints on σ_8. In practice, however, the same exponential dependence that makes cluster abundances a powerful cosmological probe also renders it susceptible to an important systematic effect, namely uncertainties in the estimated masses of clusters.

Because mass is not a direct observable, cluster masses must be determined using observable mass tracers such as X-ray emission, SZ decrements, weak lensing shear, or cluster richness (a measure of the galaxy content of the cluster). Of course, such mass estimators are noisy, meaning there can be significant scatter between the observable mass tracer and cluster mass. Since the mass function declines steeply with mass, up-scattering of low mass systems into high mass bins can result in a significant boost to the number of systems with apparently high mass (Lima & Hu 2005). If this effect is not properly modeled, the value of σ_8 derived from such a cluster sample will be overestimated.

One approach for dealing with this difficulty is to employ mass tracers that have minimal scatter, thereby reducing the impact of said scatter on the recovered halo mass function. For instance, Kravtsov et al. (2006) introduced a new X-ray mass estimator, $Y_X = M_{gas} T_X$, which in their simulations exhibits an intrinsic scatter of only $\approx 8\%$, independent of the dynamical state of the cluster. Use of a mass estimator with such low scatter should lead to improved estimates of σ_8 from X-ray cluster surveys (Pierpaoli et al. 2001; Reiprich & Böhringer 2002; Schuecker et al. 2003; Henry 2004; Stanek et al. 2006).

Such tightly-correlated mass tracers are not always available. In such cases, determination of the scatter in the mass-observable relation is critical to accurately inferring the mass function and thereby determining cosmological parameters. Of course, in practice, it is impossible to determine this scatter to arbitrary accuracy, but since the systematic boost to the mass function is proportional to the square of the scatter (Lima & Hu 2005) (i.e. the variance), even moderate constraints on the scatter can result in tight σ_8 constraints.

In this paper, we use optical and X-ray observations to constrain the scatter in the mass–richness relation for the maxBCG cluster catalog presented in Koester et al. (2007a). Specifically, we use observational constraints on the mean mass–richness relation, and on the mean and scatter of the L_X–richness relation, to convert independent estimates of the scatter in the $L_X - M$ relation into estimates of the scatter in the mass–richness relation. An interesting byproduct of our

[1] Center for Cosmology and Astro-Particle Physics (CCAPP), The Ohio State University, Columbus, OH 43210
[2] TABASGO Fellow, Physics Department, University of California at Santa Barbara, 2233B Broida Hall, Santa Barbara, CA 93106
[3] Physics Department, University of Michigan, Ann Arbor, MI 48109
[4] Astronomy Department, Universityof Michigan, AnnArbor, MI 48109
[5] Michigan Center for Theoretical Physics, Ann Arbor, MI 48109
[6] Department of Physics, The University of Chicago, Chicago, IL 60637
[7] Kavli Institute for Particle Astrophysics & Cosmology, Physics Department, and Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94305
[8] Department of Astronomy and Astrophysics, The University of Chicago, Chicago, IL 60637
[9] Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637
[10] Center for Cosmology and Particle Physics, Physics Department, New York University, New York, NY 10003
[11] Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109
[12] Fermi National Accelerator Laboratory, P.O. Box500, Batavia, IL 60510
[13] σ_8 is formally defined as the variance of the linear matter density averaged over spheres with radius $R = 8\ h^{-1}$ Mpc.

Submitted to the Astrophysical Journal
Work supported in part by US Department of Energy contract DE-AC02-76SF00515

analysis is a constraint on the correlation coefficient between mass and X-ray luminosity at fixed richness. To our knowledge, this is the first time that a correlation coefficient involving multiple cluster mass tracers has been empirically determined.

The layout of the paper is as follows. In section 1.1 we lay out the notation and definitions used throughout the paper. Section 2 presents the data sets used in our analysis. In section 3 we present a pedagogical description of our method for constraining the scatter in the richness-mass relation, while section 4 formalizes the argument. Our results are found in section 5, and we compare them to previous work in section 6. In section 7, we use our result to estimate the halo mass function in the local universe at $z = 0.23$, the median redshift of the maxBCG cluster sample, and we demonstrate that our recovered mass function is consistent with the latest cosmological constraints from WMAP (Dunkley et al. 2008). A detailed cosmological analysis of our results will be presented in a forthcoming paper (Rozo et al., in preparation). Our summary and conclusions are presented in section 8.

1.1. *Notation and Conventions*

We summarize here the notation and conventions employed in this work. Given any three cluster mass tracers (possibly including mass itself) X, Y, and Z, we make the standard assumption that the probability distribution $P(X,Y|Z)$ is a bivariate lognormal. The parameters $A_{X|Z}$, $B_{X|Z}$, and $\alpha_{X|Z}$ are defined such that

$$\langle \ln X|Z\rangle = A_{X|Z} + \alpha_{X|Z}\ln Z \quad (1)$$

$$\ln\langle X|Z\rangle = B_{X|Z} + \alpha_{X|Z}\ln Z. \quad (2)$$

Note the slopes of the mean and logarithmic mean are the same, as appropriate for a log-normal distribution. The scatter in $\ln X$ at fixed Z is denoted $\sigma_{X|Z}$, and the correlation coefficient between $\ln X$ and $\ln Y$ at fixed Z is denoted $r_{X,Y|Z}$. *We emphasize that all quoted scatters are the scatter in the natural logarithm, not in dex.* Note these parameters are simply the elements of the covariance matrix specifying the Gaussian distribution $P(\ln X, \ln Y|\ln Z)$. Under our lognormal assumption for $P(X,Y|Z)$, the parameters $A_{X|Z}$ and $B_{X|Z}$ are related via

$$B_{X|Z} = A_{X|Z} + \frac{1}{2}\sigma^2_{X|Z}. \quad (3)$$

In this work, the quantities of interest are cluster mass M, X-ray luminosity L_X, and cluster richness N. Unless otherwise specified, cluster mass is defined as M_{500c}, the mass contained within an overdensity of 500 relative to critical. L_X is the total luminosity in the rest-frame $0.5-2.0$ keV band, and N is the maxBCG richness measure N_{200}, the number of red sequence galaxies with luminosity above $0.4L_*$ within an aperture such that the mean density within said radius is, on average, $200\Omega_m^{-1}$ times the mean galaxy density assuming $\Omega_m = 0.3$. Likewise, unless otherwise stated all parameters governing the relations between M, L_X, and N assume that M is measured in units of $10^{14}\ M_\odot$, L_X is measured in units of 10^{43} ergs/s, and N is measured in "units" of 40 galaxies. For instance, including units explicitly, the mean relation between cluster mass and richness reads

$$\frac{\langle M|N\rangle}{10^{14}\ M_\odot} = \exp(B_{M|N})\left(\frac{N}{40}\right)^{\alpha_{M|N}}. \quad (4)$$

A Hubble constant parameter $h = 0.71$ is assumed throughout.[14] In addition, the weak lensing data presented in this analysis assumed a flat ΛCDM cosmology with $\Omega_m = 0.27$. The recovered mass function has the standard hubble parameter degeneracy.

2. DATA SETS

In this work we use the public maxBCG cluster catalog presented in Koester et al. (2007a), which is an optically selected volume limited catalog of close to 14,000 clusters over the redshift range $z \in [0.1, 0.3]$. These clusters were found in 7500 deg^2 of imaging data from the Sloan Digital Sky Survey (SDSS, York et al. 2000) using the maxBCG cluster finding algorithm (Koester et al. 2007b). This algorithm identifies clusters as overdensities of red sequence galaxies. All clusters are assigned a redshift based on the SDSS photometric data only, and these redshifts are known to be accurate to within a dispersion $\Delta z \approx 0.01$. Every cluster is also assigned a richness measure N_{200}, which is the number of red sequence galaxies above a luminosity cut of $0.4L_*$ and within a specified scaled aperture, centered on the Brightest Cluster Galaxy (BCG) of each cluster. Only clusters with $N_{200} \geq 10$ are included in the final catalog. Interested readers are referred to Koester et al. (2007a) and Koester et al. (2007b) for further details. In the interest of economy of notation, from now on we denote the maxBCG richness measure simply as N.

The relationship between cluster richness and various well known mass tracers has been studied in large, homogeneous samples, such as 2MASS (Dai et al. 2007) and SDSS (Becker et al. 2007; Johnston et al. 2007; Rykoff et al. 2008b; Mandelbaum et al. 2008b). Of particular interest to us are the weak lensing measurements of the mean mass as a function of richness, and the X-ray measurements of the mean and scatter of the X-ray luminosity as a function of richness. The former analysis has been carried out by Johnston et al. (2007) based on the weak lensing data presented in Sheldon et al. (2007), and independently by Mandelbaum et al. (2008a). In short, Sheldon et al. (2007) stacked maxBCG clusters within narrow richness bins, and measured the average weak lensing shear profile of the clusters. These shear profiles were turned into surface mass density contrast profiles using the redshift distribution of background sources estimated with the methods of Lima et al. (2008) and the neural net photometric redshift estimators described in Oyaizu et al. (2008). Then, Johnston et al. (2007) fit the resulting profiles using a halo model scheme to obtain tight constraints on the mean mass of maxBCG clusters for each of the richness bins under consideration. The Mandelbaum et al. (2008b) analysis is very similar in spirit to the one described above. The main differences are the way the source redshift distribution is estimated, and the details of the model fitting use to recover the masses. The differences in the results between these two analysis are discussed in appendix A.2, where we use them to set priors on the mass–richness relation.

The measurement of the mean X-ray luminosity of maxBCG clusters has been carried out by Rykoff et al. (2008b) following an approach similar to that pioneered in Dai et al. (2007). The necessary X-ray data is readily available from the ROSAT All-Sky Survey (RASS, Voges et al. 1999). In short, Rykoff et al. (2008b) stacked the RASS photon maps (Voges et al. 2001) centered on maxBCG clusters in narrow richness bins. The background subtracted stacked

[14] For other values of h, our weak lensing masses scale as $M \propto h^{-1}$ and the X-ray luminosities as $L_X \propto h^{-2}$.

photon counts within a 750 h^{-1} kpc aperture were used to estimate the mean X-ray luminosity L_X in the $0.1-2.4$ keV rest frame of the clusters. In addition, Rykoff et al. (2008b) measured the scatter in X-ray luminosity at fixed richness by individually measuring L_X for all maxBCG clusters with $N \geq 30$. It is worth noting that due to the shallowness of RASS, many of the maxBCG clusters are not X-ray luminous enough to be detected individually. However, non-detection and upper limits for L_X for individual systems were properly taken into consideration using the Bayesian approach detailed in Kelly (2007), and the recovered mean X-ray luminosity from this Baysian analysis was fully consistent with the stacked means.

In addition to the data sets above, we use the constraints on the $L_X - M$ relation from Vikhlinin et al. (2008). These constraints are based on the 400d cluster X-ray survey, a flux limited cluster survey based on ROSAT pointed observations with an effective sky coverage of 397 deg^2 (Burenin et al. 2007). Briefly, Vikhlinin et al. (2008) measured both the total soft band X-ray luminosity and the cluster mass for each cluster in the sample. X-ray luminosities are estimated from ROSAT data, and measure the luminosity in the rest-fram $0.5-2.0$ keV band, extrapolated to infinity assuming standard β profiles. Cluster masses are estimated based on the values of Y_X derived from followup Chandra observations, though they note that the results they obtain using different mass tracers such as X-ray temperature and total gas mass are very similar. The $M - Y_X$ relation is itself calibrated based on hydro-static mass estimates. Importantly, Vikhlinin et al. (2008) explicitly correct for the Malmquist bias expected for a flux limited cluster sample, so the $L_X - M$ relation they derive can be interpreted as the relation one would obtain using a mass limited cluster sample.

For this work, we have repeated the analysis in Rykoff et al. (2008b) with a slightly different definition for L_X. In particular, we measure the X-ray luminosity in the rest-frame $0.5-2.0$ keV band within a 1 h^{-1}Mpc aperture. The change in band is tailored to match the energy band used by Vikhlinin et al. (2008) , which we used to place priors on the $L_X - M$ relation. It is worth noting that Vikhlinin et al. (2008) do not use a 1 h^{-1}Mpc aperture, as we do. We have, however, carefully calibrated the scaling between our L_X definition and that of Vikhlinin et al. (2008) so as to be able to use their results in our analysis.end A detailed description of our measurements can be found in appendix A.3.

3. RELATING CLUSTER MASS, X-RAY LUMINOSITY, AND RICHNESS

The problem we are confronted with is the following: we have four pieces of observational data, namely

- The abundance of galaxy clusters as a function of richness.
- The mean relation between cluster richness and mass.
- The mean and variance of the relation between cluster richness and X-ray luminosity.
- The mean and variance of the relation between cluster X-ray luminosity and mass.

From this data, we wish to determine the scatter in mass at fixed richness for the cluster sample under consideration.

The basic idea behind our analysis is as follows. Consider the probability $P(M, L_X|N)$, which we take to be Gaussian in

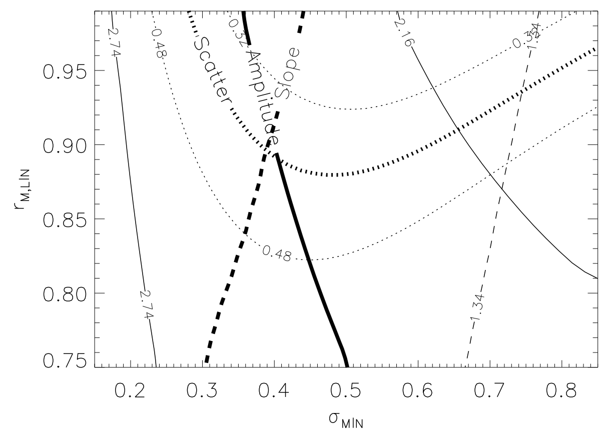

FIG. 1.— Contours of constant $L_X - M$ parameters. For each assumed value of the scatter $\sigma_{M|N}$ and correlation coefficient parameter $r_{M,L|N}$, we predict the amplitude, slope, and scatter of the $L_X - M$ relation of a mass selected sample of clusters with $M \geq 3 \times 10^{14}\ M_\odot$. Contours of constant amplitude, slope, and scatter are shown with the solid, dashed, and dotted lines respectively. The thicker lines correspond to the central values of the $L_X - M$ priors discussed in appendix A.4 and summarized in Table 1, while the the other two contours enclose the 95% confidence region for each of the parameters. The second slope contour falls to outside the region of parameter space shown in the figure. The intersection of the three separate regions correspond to acceptable values for the two unknown parameters $\sigma_{M|N}$ and $r_{M,L|N}$.

$\ln M$ and $\ln L_X$. This probability distribution is completely specified by the mean and variance of both M and L_X at fixed richness, and by the correlation coefficient between M and L_X. Of these, there are only two quantities that are not already observationally constrained: $\sigma_{M|N}$, the scatter in mass at fixed richness, and $r_{M,L|N}$, the correlation coefficient between mass and L_X at fixed richness.

Suppose now that we guessed values for these two quantities, so that the probability distribution $P(M, L_X|N)$ is fully specified. Given the abundance function $n(N)$, we can use $P(M, L_X|N)$ to randomly assign a mass and an X-ray luminosity to every cluster in the sample. We can then select a mass limited sub-sample, and measure the corresponding $L_X - M$ relation, comparing it to the $L_X - M$ measurement from Vikhlinin et al. (2008). Since the $L_X - M$ relation we predict depends on our assumptions about $P(M, L_X|N)$, there should only be a small region in parameters space where our predictions are consistent with independent observational constraints on the $L_X - M$ relation.

Figure 1 illustrates this idea. To create the figure, we have set every observed parameter of the distribution $P(M, L_X|N)$ to the central value of the priors described in appendix A and summarized in table 1. We then defined a grid in the two dimensional space spanned by $\sigma_{M|N}$ and $r_{M,L|N}$, and carried through the argument described above. The resulting predictions for the amplitude, slope, and scatter of the $L_X - M$ relation as a function of $\sigma_{M|N}$ and $r_{M,L|N}$ are shown in the figure. We plot contours of constant amplitude, slope, and scatter of the $L_X - M$ relation as solid, dashed, and dotted lines respectively. The thicker curves correspond to the central values of the priors, while thinner curves demark the corresponding 95% confidence limits. As we can see, all three contours intersect in a finite region of parameter space, indicating good agreement between our weak lensing and X-ray data, and the independent determination of the $L_X - M$ relation. Based on Figure 1, we expect a detailed analysis should constrain our parameters to $\sigma_{M|N} \approx 0.40$, and $r_{M,L|N} \approx 0.9$. The rest of this paper is simply a way of formalizing the argument described above in order to place errors on both $\sigma_{M|N}$ and $r_{M,L|N}$.

4. FORMALISM

We wish to formalize the above argument in order to place quantitative constraints on the scatter in mass at fixed richness. Details of how we go about doing so are presented below. Readers interested only in our results can move directly to section 5.

4.1. *Likelihood Model*

As we mentioned above, the key point in our analysis is our ability to compute the amplitude and slope of the mean relation $\langle \ln L_X|M\rangle$, and the scatter about this mean, as a function of our two parameters of interest: the scatter in mass at fixed richness and the correlation coefficient between M and L_X at fixed N. Let us define $\mathbf{x} = \{A_{L|M}, \alpha_{L|M}, \sigma_{L|M}\}$, and let $\mathbf{p} = \{\sigma_{M|N}, r_{M,L|N}\}$ denote our parameters of interest. Our predictions for the $L_X - M$ relation as a function of our parameters of interest can be summarized simply as $\mathbf{x}(\mathbf{p})$. Now, adopting a Bayesian framework, a set of priors on $\mathbf{x}$ is simply a probability distribution $P_{\mathbf{x}}(\mathbf{x})$. Since $\mathbf{x}$ is a function of $\mathbf{p}$, the priors immediately define a probability distribution over $\mathbf{p}$ given by

$$P(\mathbf{p}) = P_{\mathbf{x}}(\mathbf{x}(\mathbf{p}))\det(\partial\mathbf{x}/\partial\mathbf{p}). \tag{5}$$

Since we know how to compute both $P_{\mathbf{x}}(\mathbf{x})$ and $\mathbf{x}(\mathbf{p})$, we can find any confidence regions for our parameters of interest.

The problem we are confronted with, however, is slightly more complicated, in that the functions $\mathbf{x}$ depend not only on $\mathbf{p}$, but also on additional nuisance parameters $\mathbf{q}$. Indeed, our predictions for the observable parameters of the $L_X - M$ relation depend on both the abundance function of clusters and $P(M, L_X|N)$. The abundance function can be accurately described by a Schechter function (we explicitly checked a Schechter function is statistically acceptable),

$$n(N) \propto N^{-\tau}\exp(-N/N_*). \tag{6}$$

Given a Schechter fit, our prediction for the $L_X - M$ relation will also depend on the value of the parameters τ and N_*. Likewise, the distribution $P(M, L_X|N)$ also depends on the amplitude and slope of the means $\langle M|N\rangle$ and $\langle L_X|N\rangle$, as well as the scatter in L_X at fixed N. All in all, we have six additional nuisance parameters $\mathbf{q} = \{N_*, \tau, B_{M|N}, \alpha_{M|N}, A_{L|N}, \alpha_{L|N}, \sigma_{L|N}\}$. Let $\mathbf{r} = \{\mathbf{p}, \mathbf{q}\}$ denote the full set of parameters. The priors from the $L_X - M$ relation define a probability distribution over $\mathbf{r}$ given by

$$P(\mathbf{r}) = P_{\mathbf{x}}(\mathbf{x}(\mathbf{r}))\det(\partial\mathbf{x}/\partial\mathbf{r}). \tag{7}$$

Since we have a total of 8 parameters, and only three observables from the $L_X - M$ relation, it is obvious that the above likelihood function will result in large degeneracies because the parameters are under-constrained. If one has priors $P_0(\mathbf{q})$ in the nuisance parameters, however, the probability distribution $P(\mathbf{p})$ in the parameters of interest is given by

$$P(\mathbf{p}) = \int d\mathbf{q}\, P_0(\mathbf{q})P_{\mathbf{x}}(\mathbf{x}(\mathbf{p},\mathbf{q}))\det(\partial\mathbf{x}/\partial\mathbf{r}). \tag{8}$$

This equation allows us to compute $P(\mathbf{p})$, and therefore place constraints on our parameters of interest. In practice, we will ignore the determinant term in the probability distribution defined in equation 8. This is because the function $\mathbf{x}(\mathbf{r})$ is estimated using a Monte Carlo approach, implying that accurate numerical estimates of the Jacobian $\partial\mathbf{x}/\partial\mathbf{r}$ would be too computationally intensive to be performed. Fortunately, the determinant typically introduces only slight modulations of the likelihood, so we do not expect our results to be adversely affected by this.

4.2. *Implementation*

We estimate the probability distribution $P(\mathbf{p})$ using a Monte Carlo approach. Ignoring an overall normalization constant and setting $\det(\partial\mathbf{x}/\partial\mathbf{r}) = constant$, we have

$$\hat{P}(\mathbf{p}) = \frac{1}{N_{draws}}\sum_{i=1}^{N_{draws}} P_{\mathbf{x}}(\mathbf{x}(\mathbf{p},\mathbf{q}_i)) \tag{9}$$

where $\mathbf{q}_i$ for $i = 1$ through N_{draws} are random draws of the nuisance parameters $\mathbf{q}_i$, drawn from the prior distribution $P_0(\mathbf{q}_i)$. We set $N_{draws} = 3000$ as our default value (see below for further discussion).

The prior distributions for our nuisance parameters are characterized by a statistical and a systematic error. The former is modeled as Gaussian and the latter using a top-hat distribution. Thus, given a prior of the form

$$q = \bar{q} \pm \sigma_q^{stat} \pm \sigma_q^{sys}, \tag{10}$$

a random draw is obtained by setting

$$\mathbf{q}_i = \bar{\mathbf{q}} + \Delta\mathbf{q}_i^{stat} + \Delta\mathbf{q}_i^{sys} \tag{11}$$

where $\Delta\mathbf{q}_i^{stat}$ is drawn from a Gaussian of zero mean with a covariance matrix defined by the statistical errors, and $\Delta\mathbf{q}_i^{sys}$ is drawn from a top hat distribution that is non-zero only for $|\Delta q^{sys}| \leq \sigma_q^{sys}$.

The probability distribution $P_{\mathbf{x}}(\mathbf{x}(\mathbf{p},\mathbf{q}))$ used in equation 9 is the product of the likelihoods $P_x(x(\mathbf{p},\mathbf{q}))$ for each of the $L_X - M$ parameters $x \in \mathbf{x} = \{A_{L|M}, \alpha_{L|M}, \sigma_{L|M}\}$. The probability for each $L_X - M$ parameter is given by the convolution of the top-hat and Gaussian distributions defined by the statistical and systematic errors of x, so that

$$P_x(x(\mathbf{p},\mathbf{q})) = \frac{1}{4\sigma_x^{sys}}[\mathrm{erf}(x_+) - \mathrm{erf}(x_-)] \tag{12}$$

where

$$x_\pm = \frac{\pm\sigma_x^{sys} - (x(\mathbf{p},\mathbf{q}) - \bar{x})}{\sqrt{2}\sigma_x^{stat}}. \tag{13}$$

Note that the above equations are appropriate only when the various $L_X - M$ parameters are uncorrelated, so it is important to place the priors at the pivot point of the $L_X - M$ relation ($M_{pivot} = 3.9 \times 10^{14}\ M_\odot$). This explains why Table 1 quotes a prior on $A_{L|M} + 1.361\alpha_{L|M} + 1.5(\sigma_{L|M}^2 - 0.40^2)$ rather than on $A_{L|M}$ alone.

We also need to specify how the function $\mathbf{x}(\mathbf{p},\mathbf{q})$ is evaluated. We do this using a Monte Carlo approach. Given $\mathbf{p}$ and $\mathbf{q}$, we generate $N_{cl} = 10^5$ mock clusters in the richness range $N \in [10, 200]$. We then randomly draw mass and X-ray luminosity values for each of these clusters based on the distribution $P(M, L_X|N)$, and select a mass limited subsample of clusters using a mass cut $M \geq M_{min}$ with $M_{min} = 3 \times 10^{14} M_\odot$ (the reason for this particular value is explained below). Using a least squares fitting routine, we find the best fit line between $\ln L_X$ and $\ln M$. This defines both $A_{L|M}(\mathbf{p},\mathbf{q})$ and $\alpha_{L|M}(\mathbf{p},\mathbf{q})$. The scatter $\sigma_{L|M}(\mathbf{p},\mathbf{q})$ is defined as the root mean square fluctuation about the best fit line.

Using equation 9 and the function $\mathbf{x}(\mathbf{p},\mathbf{q})$ defined above, we evaluate the probability distribution $P(\mathbf{p})$ along a grid of points in $\sigma_{M|N} \in [0.2, 0.85]$ and $r \in [0.75, 1.0]$ with 25 grid points per axis. A full run of our code then requires we perform 25^2 Monte Carlo integrals with $N_{draws} = 3000$ points in each integration. Each draw also requires us to evaluate the function $\mathbf{x}(\mathbf{p},\mathbf{q})$, which in turn requires generating a mock

TABLE 1
SCALING RELATION AND CLUSTER ABUNDANCE PRIORS

Parameter	Prior
$\ln N_*$	3.66 ± 0.10 $(stat) \pm 0.01$ (sys)
τ	2.61 ± 0.06 $(stat) \pm 0.05$ (sys)
$B_{M\|N}$	0.95 ± 0.07 $(stat) \pm 0.10$ (sys)
$\alpha_{M\|N}$	1.06 ± 0.08 $(stat) \pm 0.08$ (sys)
$B_{L\|N}$	1.91 ± 0.04 $(stat) \pm 0.09$ (sys)
$\alpha_{L\|N}$	1.63 ± 0.06 $(stat) \pm 0.05$ (sys)
$\sigma_{L\|N}$	0.83 ± 0.03 $(stat) \pm 0.10$ (sys)
$A_{L\|M} + 1.361\alpha_{L\|M} + 1.5(\sigma^2_{L\|M} - 0.40^2)$	2.45 ± 0.08 $(stat) \pm 0.23$ (sys)
$\alpha_{L\|M}$	1.61 ± 0.14 $(stat)$
$\sigma_{L\|M}$	0.40 ± 0.04 $(stat)$

Priors on the abundance function parameters (N_* and τ), as well as those from the $M-N$ and L_X-N relations are not taken directly from any single work in the literature, but are discussed in detail in Appendix A. Priors on the L_X-M relation are taken from Vikhlinin et al. (2008). Overall, we believe these priors are fair, that is, they are neither overly optimistic nor overly pessimistic.

catalog with $N_{cl} = 10^5$ clusters, so the procedure as a whole is computationally expensive. To increase computational efficiency, for each Monte Carlo evaluation of $P(\mathbf{p})$ we generate a single cluster catalog that is used to estimate the likelihood at every grid point. This correlates the values of $\hat{P}$ along our grid, but does not otherwise adversely affect our results.

Our Monte Carlo approach requires that both the number of clusters in the random catalogs N_{cl} and the number of times the likelihood function is evaluated N_{draws} is sufficiently large to achieve convergence. Our default values for N_{cl} and N_{draws} were selected to ensure the recovered likelihood is accurate to within a dispersion of $\sim 1-2\%$ inside high likelihood regions. The error in the recovered likelihood increases with decreasing likelihood, but even in the tails of the distributions our estimates are accurate to about 10%. This was explicitly tested by running a coarse grid with our default values for N_{draws} and N_{cl}, and by repeating the analysis with both of these parameters increased by a factor of two.[15]

Finally, we emphasize that it is necessary to explicitly check whether our results are sensitive to the $N \geq 10$ cut applied to the maxBCG clusters sample. In particular, when selecting a mass limited subsample of clusters, we need to ensure that the mass limit M_{min} be sufficiently large that the number of clusters with $N \leq 10$ and $M \geq M_{min}$ is insignificant. We have explicitly checked that for our adopted low mass cut $M_{min} \geq 3\times 10^{14}\ M_\odot$ our results are robust to the richness cut $N \geq 10$ by repeating the analysis in a coarse grid using an $N \geq 8$ richness cut instead. We find that the likelihood estimates in both cases are in agreement to within the expected accuracy of our Monte Carlo approach.

4.3. *Priors*

The priors used in our analysis are summarized in Table 1. We follow the notation

$$q = \bar{q} \pm \sigma_q^{stat}\ (stat) \pm \sigma_q^{sys}\ (sys) \qquad (14)$$

where $\bar{q}$ is the central value, σ_q^{stat} is the 1σ statistical error on the parameter q marginalized over all other parameters, and σ_q^{sys} is the systematic error. In all cases, we model statistical errors as Gaussian, and we include known covariances between different parameters. Systematic errors are assumed to follow top-hat distributions, and the final prior distribution is given by the convolution of these two functions.

We believe that the priors contained in table 1 are fair, that is, they are neither overly aggressive nor overly conservative. A detailed discussion of our priors can be found in appendix A.

[15] It is worth noting that in order to create Figure 1, one needs to generate cluster catalogs with $N_{cl} \gtrsim 10^7$ clusters in order for the contours to appear smooth by eye. However, $N_{cl} = 10^5$ is a sufficient number of clusters for our analysis, since we only require that the noise in the likelihood be much smaller than the width of the priors. Since the latter are quite wide, even relatively noisy estimates of the L_X-M relation are sufficient for constraining the marginalized distribution.

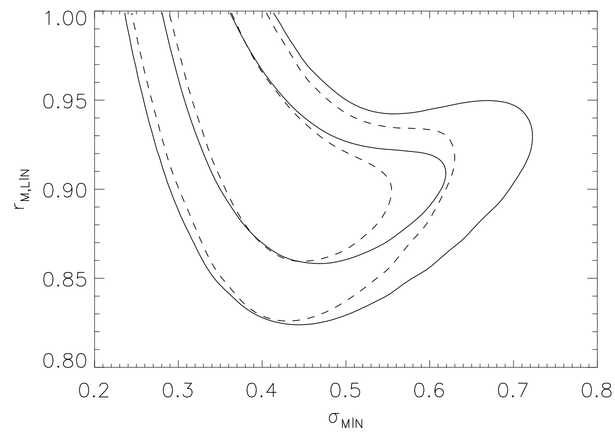

FIG. 2.— 68% and 95% confidence contours for $\sigma_{M|N}$ and $r_{L,M|N}$. Solid lines show the results of our analysis. We find that X-ray luminosity and mass are correlated at fixed richness. The breadth of the degeneracy region shown above is almost exclusively due to uncertainties in the L_X-M relation parameters. Dashed contours demonstrate how our results would improve if the L_X-M amplitude and slope were known to within an accuracy of $\Delta A_{L|M} = \Delta\alpha_{L|M} = 0.05$.

5. RESULTS

Figure 2 shows the 68% and 95% probability contours for the parameters $\sigma_{M|N}$ and $r_{M,L|N}$. The likelihood peak occurs at $\sigma_{M|N} = 0.46$ and $r_{M,L|N} = 0.90$. The marginalized means are $\langle\sigma_{M|N}\rangle = 0.45$ and $\langle r_{M,L|N}\rangle = 0.91$.

We wish to determine whether the breadth of the likelihood region in Figure 2 is limited by uncertainties in the scaling relations of maxBCG clusters, or by uncertainties in the L_X-M relation. To do so, we repeat our analysis with two new sets of priors: for the first, we use a tight 0.05 statistical prior on all nuisance parameters, but let the L_X-M parameters float. The second set of priors uses a tight 0.05 prior on each of the L_X-M parameters, but floats all other nuisance parameters with the original priors. We find that using tight priors on our nuisance parameters has negligible impact on the likelihood regions recovered from our analysis. On the other hand, the confidence regions obtained with the tight L_X-M priors, shown in Figure 2 as dashed curves, are tighter than those derived from our original analysis. Thus, the dominant source of error in our analysis is the uncertainty in the values of the L_X-M parameters. This can be easily understood based on Figure 1. We can see from the figure that the uncertainty in $r_{M,L|N}$ is largely due to the prior on the scatter in L_X at fixed M, which is already tight and thus does not change between our fiducial prior and our tight priors. On the other hand, we can see that both the amplitude and slope priors cut-off regions with high scatter. Tightening these priors excludes a larger section of parameter space, and results in the tighter contours observed in Figure 2.

Figure 3 shows the marginalized probability distributions for $\sigma_{M|N}$ and $r_{M,L|N}$. The solid curves correspond to our original analysis, while the dashed curves illustrate the results one expects assuming our hypothetical tight priors for the $L_X - M$ relation parameters. We find that the logarithmic scatter in mass at fixed richness and the correlation coefficient between $\ln M$ and $\ln L_X$ are

$$\sigma_{M|N} = 0.45^{+0.20}_{-0.18}\ (95\%\ \mathrm{CL}) \tag{15}$$
$$r_{L,M|N} \geq 0.85\ (95\%\ \mathrm{CL}). \tag{16}$$

Assuming our hypothetical tight $L_X - M$ priors, the constraints become $\sigma_{M|N} = 0.42^{+0.07}_{-0.09}$ and $r_{L,M|N} \geq 0.85$ (95% CL). We emphasize that these latter constraints are only meant as a guide to the accuracy one could achieve with this method if the $L_X - M$ relation were known to about 5% accuracy.

It is evident from our results that cluster richness is not as effective a mass tracer as X-ray derived masses. Indeed, even total (i.e. not core-core excluded) X-ray luminosity is a more faithful mass tracer than the adopted richness measure of the maxBCG catalog, as demonstrated both by the smaller scatter and the very large correlation coefficient. Note that the latter indicates that, at fixed richness, over-luminous clusters are almost guaranteed to also be more massive than average. This is an important result which forms the basis for a concurrant paper in which we improve our richness estimates by demanding tighter correlations in the L_X–richness relation (Rozo et al. 2008).

6. COMPARISON TO OTHER WORK

There are not many previous results against which our measurements of scatter in mass at fixed richness may be compared. One possible reference point is the upper limit based on the error bar in the weak lensing mass estimates of Johnston et al. (2007). More specifically, assuming that the error in $\langle M|N\rangle$ is entirely due to the intrinsic scatter in M at fixed N, it follows that the error in the mass is simply $\Delta M/\langle M|N\rangle \approx \Delta \ln M = \sigma_{M|N}/\sqrt{n(N)}$ where ΔM is the observed error and $n(N)$ is the number of clusters with richness N. For the richest bin, which provides the tightest constraint, Johnston et al. (2007) find $\langle M\rangle = (8.1 \pm 1.3) \times 10^{14}\ M_\odot$. The bin contains $n = 47$ clusters, so an upper limit to the scatter in mass at fixed richness is $\sigma_{M|N} \leq \sqrt{n}(\Delta M/\langle M\rangle) = 1.10$. Figure 3 shows that our results easily satisfy this upper limit on the scatter.

The only other measurement of the scatter in mass at fixed richness for maxBCG clusters is that found in Becker et al. (2007). These scatter estimates are obtained as follows: first, Becker et al. (2007) select all maxBCG clusters whose central galaxy has a spectroscopic redshift. They then bin the clusters in richness, and compute the velocity relative to the BCG of every galaxy member with spectroscopic data. The recovered velocity distribution of galaxies is found to be non-Gaussian. Assuming that the velocity distribution of galaxies of halos of fixed mass is exactly Gaussian, and that the observed non-Gaussianity is entirely due to mass-mixing within a richness bin, Becker et al. (2007) estimate the scatter in mass at fixed richness based on the observed non-Gaussianity of the velocity distribution.

An updated version of the results from Becker et al. (2007) can be seen in Figure 4. The only difference between this plot and the corresponding figure in Becker et al. (2007) is that here we have made used of the additional spectroscopic data from the SDSS Data Release 6 (Adelman-McCarthy et al. 2008), which results in tighter error bars. Also shown in the figure as a horizontal band is the 95% confidence region from

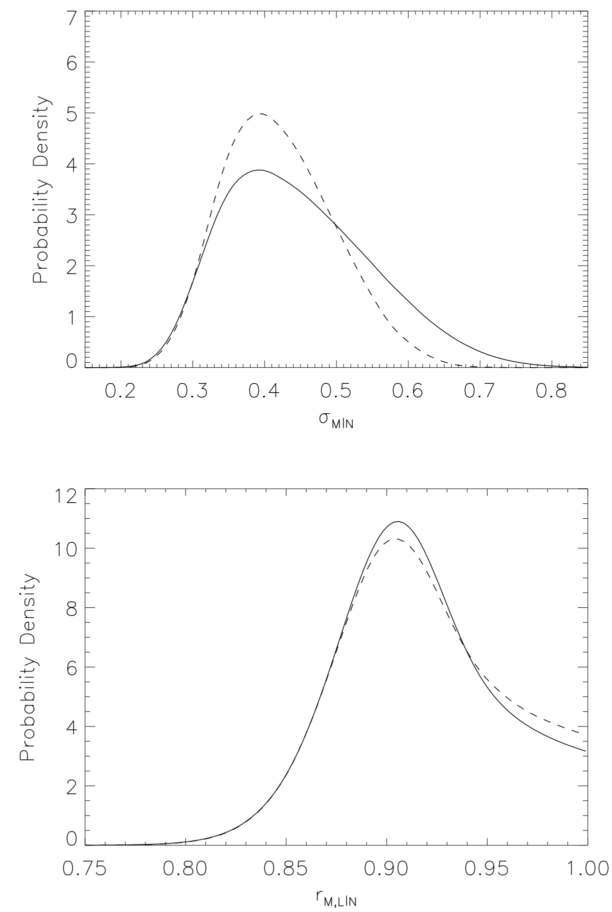

FIG. 3.— Likelihood distributions for $\sigma_{M|N}$ and $r_{M,L|N}$. The distributions are marginalized over all other parameters. Solid lines are the results of our analysis, while dashed lines are the results obtained assuming tight priors on the $L_X - M$ parameters Note the latter set of curves are presented only to give a sense of how our result would improve with better understanding of the $L_X - M$ relation.

our analysis. As we can see, our scatter estimate appears to be systematically lower than that of Becker et al. (2007), a discrepancy first noted in Rykoff et al. (2008a, more on the relation between our work and theirs below).

Such a bias is not entirely unexpected, as we now know that a significant fraction of cluster have their BCGs missidentified, a problem that was not yet known – and was therefore unaccounted for – at the time the Becker et al. (2007) results came out. To get a better understanding of how our results and those of Becker et al. (2007) compare, we can use our results along with the miscentering probability model from Johnston et al. (2007) to predict the scatter that Becker et al. (2007) observed given this miscentering systematic. We proceed as follows. First, we use our best fit model for the abundance distribution to generate a mock catalog with 2×10^5 clusters with $N \geq 10$. Each of these clusters is assigned a mass by drawing from the $P(M|N)$ distribution defined by the values of $\sigma_{M|N}$ corresponding to the two 95% confidence limits on $\sigma_{M|N}$. These assigned masses are then turned into velocity dispersions using the scaling relation from Evrard et al. (2008).

At this point, we have a cluster catalog where each cluster has a richness and a velocity dispersion. If a cluster is miscentered, we expect that in most cases the new center will be a cluster galaxy. Assuming this is the case, and that BCGs are

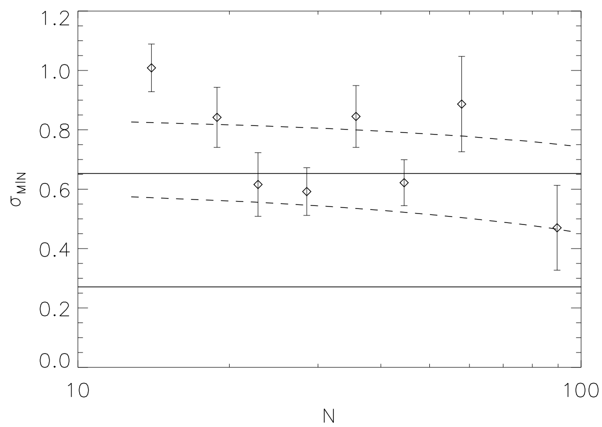

FIG. 4.— Comparison of the scatter in mass at fixed richness estimated in this work (solid band) and that of Becker et al. (2007) (diamonds with error bars). The dashed band shows how the scatter we measured is expected to be affected by miscentering, which allows us better compare our results to those of Becker et al. (2007). We find that, once miscentering is properly taken into account, the two results appear to be in reasonable agreement.

at rest at the center of a cluster, the velocity dispersion of cluster galaxies relative to random satellites will be a factor of $\sqrt{2}$ high than relative to the BCG. Using the miscentering model described in Johnston et al. (2007) for $p(N)$, the probability that a cluster of richness N be correctly centered, we randomly label clusters as properly centered or miscentered, and boost their "observed" velocity dispersion for those clusters labeled as miscentered by the expected amount. The clusters are assigned a new mass based on their "observed" velocity dispersions, and the corresponding scatter in the $M-N$ relation is estimated. We repeat this procedure 10^3 times in order to compute the mean systematic correction due to miscentering.

Our predictions for the scatter values observed by Becker et al. (2007) are shown in Figure 4 with dashed lines, and correspond to the 95% confidence interval from our analysis. We see that miscentering introduces a richness dependent correction that boosts the scatter in the recovered velocity dispersion and places it in significantly better agreement with the data from Becker et al. (2007).

The agreement with the Becker et al. (2007) data is an interesting result. Perhaps the single most difficult systematic effect that had to be addressed in the Becker et al. (2007) analysis is the validity of the assumption that non-Gaussianities in the velocity distribution of stacked clusters are entirely due to mass-mixing is a valid. The reasonable agreement between our results and those of Becker et al. (2007) suggests that their assumption is indeed justified, though a robust conclusion will have to wait until a more detailed analysis is performed, especially given the possibility of velocity bias of the galaxy population (i.e. if satellite galaxies have a velocity dispersion different from that of the dark matter).

The analysis in this work is also very closely related to that of Rykoff et al. (2008a). Rykoff et al. (2008a) sought to constrain the L_X-M relation of clusters by fitting the scaling of $\langle L_X|N\rangle$ with $\langle M|N\rangle$. However, as recognized in Rykoff et al. (2008a), in order to fully interpret their result in terms of the traditional definition of the L_X-M relation, i.e. the mean X-ray luminosity at fixed mass, one needs to know both the scatter in mass at fixed richness, and the corresponding correlation coefficient with L_X. Given that these two quantities are unknown, but that the L_X-M relation is already constrained from X-ray surveys, it seems reasonable to suggest that a better use of the lensing and X-ray data of maxBCG clusters is to use our knowledge of the L_X-M relation to constrain the scatter in mass at fixed richness and the corresponding correlation coefficient, as was done in this work.

Our work differs from the ideas presented in Rykoff et al. (2008a) in another significant way. While our analysis employs only $P(L_X,M|N)$ and $n(N)$, Rykoff et al. (2008a) used the halo mass function dn/dM and the probability distribution $P(L_X,N|M)$ to interpret their measurements. This has the important drawback that in doing so, one needs to assume a cosmological model in order to compute the halo mass function, rendering their interpretation cosmology dependent. By focusing on the quantities that are directly observable, i.e. $n(N)$ and $P(L_X,M|N)$, we are able to avoid this difficulty. The price we pay for this is that rather than constraining the scatter in richness at fixed mass, which is the more directly relevant quantity from a cosmological perspective, we constrain instead the scatter in mass at fixed richness. While this makes implementing such a constraint a little more cumbersome in a cosmological analysis, the fact that the constraint itself is cosmology independent is obviously of paramount importance.

7. COSMOLOGICAL CONSEQUENCES

As mentioned in the introduction, to obtain an unbiased estimate of the halo mass function based on the observed cluster richness function requires that we understand the scatter between cluster richness and halo mass. Given our lognormal assumption, and the fact that the mean mass–richness relation is already known from weak lensing, our measurement of the scatter in this scaling relation fully determines the probability distribution $P(M|N)$. Thus, we are now in a position to determine the halo mass function of the local universe with the maxBCG cluster catalog.

Let us define then $n_i = n(M_i)$ as the number of halos within a logarithmic mass bin of width $\Delta \ln M$ centered about M_i,

$$n_i = \left.\frac{dn}{d\ln M}\right|_{M_i} \Delta \ln M. \qquad (17)$$

Given our cluster catalog and $P(M|N)$, we construct an estimator $\hat{n}_i$ for n_i by randomly drawing a mass from $P(M|N)$ for each halo in the cluster catalog, and then counting the number of halos within the logarithmic mass bin centered about M_i. Note that since the mass of each cluster is a random variable, our mass function estimator $\hat{n}_i$ is itself a random variable. The mean and correlation matrix of $\hat{n}_i$ can easily be obtained by making multiple realizations of $\hat{n}_i$, and averaging the resulting mass functions.

In practice, we also need to marginalize our results over uncertainties in $P(M|N)$ and over uncertainties in the richness function $n(N)$. To do so, we randomly draw the parameters $\mathbf{x} = \{B_{M|N}, \alpha_{M|N}, \sigma_{M|N}\}$, and then resample of the cluster richness function to obtain a new estimate of n_i. The whole procedure is iterated 10^5 times, and the mean and covariance matrix of the number counts in each of our logarithmic mass bins is computed.[16]

Figure 5 shows the mass function recovered through our analysis. To turn our number counts into a density, we assumed a WMAP5 cosmology (Dunkley et al. 2008), with $\Omega_m = 0.27$ and $h = 0.72$, and a photometric redshift error

[16] We again checked explicitly that the mass cut $M_{min} = 3\times10^{14}\ M_\odot$ is large enough for our results to be insensitive to the maxBCG richness cut $N \geq 10$.

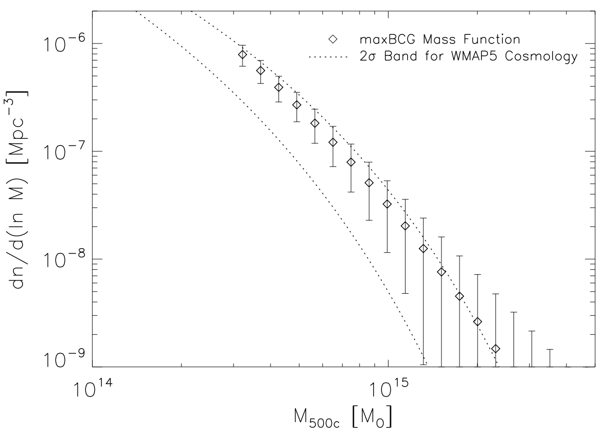

FIG. 5.— The maxBCG mass function. Cluster counts were converted to densities assuming $\Omega_m = 0.27$ and $h = 0.71$, the same cosmology assumed in the lensing measurements (Johnston et al. 2007). The error bars shown are due to the scatter in the mass–richness relation, and are strongly correlated. For comparison, we have also plotted the Tinker et al. (2008) mass function corresponding to the WMAP5 95% confidence region for σ_8, $0.724 \leq \sigma_8 \leq 0.868$. All other parameters are held fixed to the central values reported in Dunkley et al. (2008). Our data are consistent with the WMAP5 results, though they might suggest a slightly higher power spectrum normalization.

$\Delta z = 0.01$ (used for computing the effective volume of the sample). The diamonds correspond to our estimated means, and the error bars are the square root of the diagonal elements of the correlation matrix. We emphasize that the error bars are heavily correlated. The mean and covariance matrix of the recovered halo mass function can be found in Appendix B.

Also shown in Figure 5 with dotted lines are the halo mass functions at $z = 0.23$ predicted by WMAP5 assuming the Tinker et al. (2008) mass function. For both curves, we set all cosmological parameters to the central values reported in Dunkley et al. (2008), except for σ_8, which is set to $\sigma_8 = 0.868$ for the upper curve and $\sigma_8 = 0.724$ for the lower curve. These two values define the 95% confidence interval for σ_8 in Dunkley et al. (2008). As we can see, the mass function recovered from our analysis is fully consistent with the WMAP5 cosmology, though it seems to push for values of σ_8 on the high end of their allowed region. A detailed cosmological analysis of our data will be presented in a subsequent paper (Rozo et al, in preparation).

8. SUMMARY AND CONCLUSIONS

We have shown that by combining the information in the maxBCG richness function, the mean richness-mass relation, the mean and scatter of the L_X–richness relation, and the mean and scatter of the $L_X - M$ relation, we can constrain both the scatter in mass at fixed richness for maxBCG clusters, as well as the correlation coefficient between mass and L_X at fixed richness. We find

$$\sigma_{M|N} = 0.45^{+0.20}_{-0.18} \text{ (95\% CL)} \qquad (18)$$

$$r_{L,M|N} \geq 0.85 \text{ (95\% CL).} \qquad (19)$$

These constraints are dominated by uncertainties in the $L_X - M$ relation, and can be significantly tightened if our understanding of the $L_X - M$ relation improves. We also found our results are consistent with those presented in Becker et al. (2007) once miscentering of maxBCG clusters is taken into account.

Our lower limit on the correlation between M and L_X at fixed richness constitutes the first observational constraint on a correlation coefficient involving two different halo mass tracers. Note that the large correlation between L_X and M implies that L_X - even without core exclusion - is a significantly better mass tracer than the maxBCG richness estimator (i.e. at fixed richness, over-luminous cluster are nearly always more massive). This is an important result, which we use in a concurrent paper to help us define new richness estimators that are better correlated with cluster mass (Rozo et al. 2008).

Using our results, and assuming $\Omega_m = 0.27$ and $h = 0.71$, we have estimated the halo mass function at $z = 0.23$, corresponding to the median redshift of the cluster sample. We find that our recovered mass function is in good agreement with the mass function predicted by Tinker et al. (2008) for the WMAP5 cosmology (Dunkley et al. 2008). A detailed cosmological analysis will be presented in a forthcoming paper (Rozo et al, in preparation).

Our work sheds new light on the interrelationship of bulk properties of massive halos. We have used weak lensing, X-ray luminosities, and optical richness estimates to constrain the scatter in the richness-mass relation, which can lead to improved cosmological constraints. In principle, one could also turn this question around, and, assuming cosmology, we could constrain the scatter in the richness-mass relation, which would then allow us to place constraints on the amplitude, slope, and scatter of the $L_X - M$ relation. Such an analysis would be interesting in that, by doing so, one could compare the predicted amplitude of the $L_X - M$ relation to that derived from hydrostatic mass estimates, thereby directly probing the amount of non-thermal pressure support in galaxy clusters. Note that even though this question can also be directly addressed by comparing weak lensing and X-ray mass estimates of individual clusters, the analysis suggested here would benefit from having small uncertainties, whereas projection effects result in rather noisy weak lensing mass estimates for individual systems.

The authors would like to thank Alexey Vikhlinin for providing them with the full covariance of the $L_X - M$ parameters from their analysis of the 400d cluster sample. ER thanks David Weinberg and Chris Kochanek for useful discussions and their careful reading of the manuscript. ESR would like to thank the TABASGO foundation. TM and JH gratefully acknowledge support from NSF grant AST 0807304 and DoE Grant DE-FG02-95ER40899. AE gratefully acknowledges support from NSF grant AST-0708150. RHW was supported in part by the U.S. Department of Energy under contract number DE-AC02-76SF00515 and by a Terman Fellowship at Stanford University. This project was made possible by workshop support from the Michigan Center for Theoretical Physics.

REFERENCES

Adelman-McCarthy, J. K. et al. 2008, ApJS, 175, 297
Bardeau, S., Kneib, J.-P., Czoske, O., Soucail, G., Smail, I., Ebeling, H., & Smith, G. P. 2005, A&A, 434, 433

Bardeau, S., Soucail, G., Kneib, J.-P., Czoske, O., Ebeling, H., Hudelot, P., Smail, I., & Smith, G. P. 2007a, A&A, 470, 449
—. 2007b, A&A, 470, 449
Becker, M. R. et al. 2007, ApJ, 669, 905
Burenin, R. A., Vikhlinin, A., Hornstrup, A., Ebeling, H., Quintana, H., & Mescheryakov, A. 2007, ApJS, 172, 561
Dahle, H. 2006, ApJ, 653, 954
Dai, X., Kochanek, C. S., & Morgan, N. D. 2007, ApJ, 658, 917
Dunkley, J. et al. 2008, ArXiv e-prints, 803, astro-ph/0803.0586
Evrard, A. E. 1990, ApJ, 363, 349
Evrard, A. E. et al. 2008, ApJ, 672, 122
Frenk, C. S., White, S. D. M., Efstathiou, G., & Davis, M. 1990, ApJ, 351, 10
Gladders, M. D., Yee, H. K. C., Majumdar, S., Barrientos, L. F., Hoekstra, H., Hall, P. B., & Infante, L. 2007, ApJ, 655, 128
Haiman, Z., Mohr, J. J., & Holder, G. P. 2001, ApJ, 553, 545
Henry, J. P. 2004, ApJ, 609, 603
Henry, J. P., & Arnaud, K. A. 1991, ApJ, 372, 410
Hoekstra, H. 2007, MNRAS, 379, 317
Holder, G., Haiman, Z., & Mohr, J. J. 2001, ApJ, 560, L111
Hu, W., & Kravtsov, A. V. 2003, ApJ, 584, 702
Johnston, D. E. et al. 2007, ArXiv e-prints, 709, astro-ph/0709.1159
Kelly, B. C. 2007, ApJ, 665, 1489
Koester, B. P. et al. 2007a, ApJ, 660, 239
—. 2007b, ApJ, 660, 221
Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, ApJ, 650, 128
Lima, M., Cunha, C. E., Oyaizu, H., Frieman, J., Lin, H., & Sheldon, E. S. 2008, ArXiv e-prints, 801, astro-ph/0801.3822
Lima, M., & Hu, W. 2004, Phys. Rev. D, 70, 043504
—. 2005, Phys. Rev. D, 72, 043006
Mahdavi, A., Hoekstra, H., Babul, A., & Henry, J. P. 2008, MNRAS, 384, 1567
Mandelbaum, R., Seljak, U., & Hirata, C. M. 2008a, ArXiv e-prints, 805, astro-ph/0805.2552
Mandelbaum, R. et al. 2008b, MNRAS, 386, 781
Maughan, B. J. 2007, ApJ, 668, 772
Nagai, D., Vikhlinin, A., & Kravtsov, A. V. 2007, ApJ, 655, 98
Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., Frieman, J., & Sheldon, E. S. 2008, ApJ, 674, 768
Pierpaoli, E., Scott, D., & White, M. 2001, MNRAS, 325, 77
Rasia, E. et al. 2006, MNRAS, 369, 2013
Reiprich, T. H., & Böhringer, H. 2002, ApJ, 567, 716
Rozo, E., Dodelson, S., & Frieman, J. A. 2004, Phys. Rev. D, 70, 083008
Rozo, E., Wechsler, R. H., Koester, B. P., Evrard, A. E., & McKay, T. A. 2007a, ArXiv Astrophysics e-prints, astro-ph/0703574
Rozo, E. et al. 2007b, ArXiv Astrophysics e-prints, astro-ph/0703571
Rozo, E., Zentner, A. R., Bertone, G., & Chen, J. 2006, ApJ, 639, 573
Rozo, E., et al. 2008, ArXiv e-prints, 809, astro-ph/0809.2797
Rykoff, E. S. et al. 2008a, MNRAS, 387, L28
—. 2008b, ApJ, 675, 1106
Schuecker, P., Böhringer, H., Collins, C. A., & Guzzo, L. 2003, A&A, 398, 867
Sheldon, E. S. et al. 2007, ArXiv e-prints, 709, astro-ph/0709.1153
Stanek, R., Evrard, A. E., Böhringer, H., Schuecker, P., & Nord, B. 2006, ApJ, 648, 956
Tinker, J. L., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M. S., Yepes, G., Gottlober, S., & Holz, D. E. 2008, ArXiv e-prints, 803, astro-ph/0803.2706
Vikhlinin, A. et al. 2008, ArXiv e-prints, 805, astro-ph/0805.2207
Voges, W. et al. 1999, A&A, 349, 389
Voges, W., Boller, T., Englhauser, J., Freyberg, M., & Supper, R. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 225, Virtual Observatories of the Future, ed. R. J. Brunner, S. G. Djorgovski, & A. S. Szalay, 234–+
York, D. G., Adelman, J., Anderson, J. E., Anderson, S. F., Annis, J., & the SDSS collaboration. 2000, AJ, 120, 1579
Zhang, Y.-Y., Finoguenov, A., Böhringer, H., Kneib, J.-P., Smith, G. P., Czoske, O., & Soucail, G. 2007, A&A, 467, 437
Zhang, Y.-Y., Finoguenov, A., Böhringer, H., Kneib, J.-P., Smith, G. P., Kneissl, R., Okabe, N., & Dahle, H. 2008, A&A, 482, 451

APPENDIX

PRIORS

Abundance Priors

Our estimates of the $L_X - M$ parameters depend on the abundance function of maxBCG clusters, which is observationally determined, but not known to infinite precision. Here, we fit the observed abundance function using a Schechter function, such that the mean number of clusters μ of richness N is

$$\mu(N) = n_0(N/40)^{-\tau}\exp(-N/N_*). \tag{A1}$$

The amplitude n_0 is chosen such that the total number of clusters exactly equals the observed number of clusters. We set this normalization condition because we are interested only in the shape of the richness function, and not in its amplitude.

The fits are done by maximizing the likelihood of the observed distribution, binned in bins of width $\Delta N = 1$. We assume that the probability distribution of observed n clusters in a bin of richness N is Poisson, with

$$P(n) = \frac{\mu(N)^n \exp(-\mu(N))}{n!}. \tag{A2}$$

For numerical purposes, we cut the distribution at $N_{max} = 300$, which is sufficiently large to not affect our fits. We emphasize that we use the above likelihood only to define estimators for N_* and τ, since, as discussed below, both goodness of fit and errors in the parameter estimation are obtained through Monte Carlo simulation.

The richness distribution is fit over the range $N \geq 10$ by maximizing the log-likelihood function using an amoeba routine. To estimate our errors, we follow a Monte Carlo approach and resample the observed richness function 10^4 times. We find that the parameters N_* and τ are significantly correlated, with the probability distribution being Gaussian in τ and $\ln N_*$. The best fit parameters are

$$\langle \ln N_* \rangle = 3.66 \pm 0.10 \tag{A3}$$

$$\langle \tau \rangle = 2.61 \pm 0.06 \tag{A4}$$

with a correlation coefficient

$$r_{N_*,\tau} = 0.94. \tag{A5}$$

To assess goodness of fit, we generate 10^4 mock catalogs with as many clusters as the real data from the probability distribution specified by $\langle \ln N_* \rangle$ and $\langle \tau \rangle$. We compute the likelihood for each of these mock catalogs, and compare the corresponding likelihood distribution to that observed in the real data. We find that our fit is statistically acceptable.

The most significant systematic error affecting our measurements of the shape of the richness function are completeness and purity variations in the cluster catalog. Rozo et al. (2007a) have shown that the maxBCG catalog is over 90% pure and complete for $N \geq 10$. Here, we take a conservative approach, and consider the change in the best fit parameters assuming the observed counts are rescaled by a completeness/purity correction factor λ given by

$$\lambda = \min\{0.9 + 0.1\ln(N/10.0)/\ln(10.0)\}. \tag{A6}$$

This corresponds to a 10% decrease in the observed counts at $N = 10$ while holding the counts at $N = 100$ constant. Upon refitting the data after this correction we find systematic offsets

$$(\Delta \ln N_*)_{sys} = 0.01 \tag{A7}$$

$$(\Delta\tau)_{sys} = 0.05 \tag{A8}$$

which we adopt as our systematic error. Note the systematic offsets are allowed to be both positive and negative, since the correction multiplier λ above could easily be larger than unity rather than smaller than unity.

$M-N$ Priors

Our priors on the $M-N$ relation are based on the results presented in Johnston et al. (2007), Mandelbaum et al. (2008b), and Mandelbaum et al. (2008a). To assign our priors, we first compare the results of these two works as a means of assessing systematic uncertainties in the mass parameters. We then focus exclusively on the Johnston et al. (2007) results to place our final priors on the $M-N$ relation. The latter choice reflects the fact that Johnston et al. (2007) report weak lensing mass estimates for several mass definitions, among them M_{500c}, the relevant quantity in the $L_X - M$ relation of Vikhlinin et al. (2008).

Let us then begin by discussing the Johnston et al. (2007) results first. While Johnston et al. (2007) quote a power-law fit for the mean mass at fixed $\langle M|N\rangle$, this fit is based a non-public version of the maxBCG catalog that extends to a richness of $N = 3$ (the catalog for clusters with $N < 10$ is not public). Since the maxBCG catalog is only known to be highly complete and pure in the range $N \geq 10$, we have refit the Johnston et al. (2007) masses restricting ourselves to the range $N \geq 9$. This slightly lower cut is necessary due to the richness binning in Johnston et al. (2007). We find that the mass M_{180b} within a 180 overdensity threshold relative to mean matter density is

$$\frac{\langle M_{180b}|N\rangle}{10^{14}\ h^{-1}M_\odot} = \exp(0.25 \pm 0.07)(N/20)^{1.18\pm0.09} \tag{A9}$$

with a correlation coefficient $r = -0.43$ between the amplitude and slope parameters.

Mandelbaum et al. (2008a) preformed a similar but independent weak lensing analysis of the maxBCG clusters, though using M_{200b} as their mass variable. They find

$$\frac{\langle M_{200b}|N\rangle}{10^{14}\ h^{-1}M_\odot} = \exp(0.45 \pm 0.08)(N/20)^{1.15\pm0.14} \tag{A10}$$

To compare against the Johnston et al. (2007) values, we use the Hu & Kravtsov (2003) mass conversion formulae to find an approximate power law relation between M_{200b} and M_{180b} over the range $5\times10^{14}\ M_\odot \leq M_{200b} \leq 10^{15}\ M_\odot$. We find $M_{180b} = 1.022M_{200b}$, which is only a 2% correction. Applying this correction, we find that the corresponding $M-N$ parameters from Mandelbaum et al. (2008a) are

$$\frac{\langle M_{180b}|N\rangle}{10^{14}\ h^{-1}M_\odot} = \exp(0.47 \pm 0.07)(N/20)^{1.15\pm0.14} \tag{A11}$$

We find that the slopes of the Johnston et al. (2007) and Mandelbaum et al. (2008a) results are nearly identical, but that the masses of Mandelbaum et al. (2008a) are systematically higher by $\approx 25\%$. This difference can be traced back to how the lensing critical surface density for each of the two works is estimated.

In general, lensing masses are proportional to the quantity $1/\left\langle\Sigma_{crit}^{-1}\right\rangle$, where Σ_{crit} is the lensing critical surface density, and the average is to be computed over the source redshift distribution. Given multi-band photometric data $\mathbf{m}$ for each galaxy, one way to compute $\left\langle\Sigma_{crit}^{-1}\right\rangle$ is to use a photometric redshift estimator $z_{photo}(\mathbf{m})$, and then assume that the true source redshift distribution is identical to the photometric redshift distribution. Mandelbaum et al. (2008b) have shown that such a simple approach typically results in biased lensing mass estimates, but they also demonstrate that it is possible to achieve unbiased results using the probability distribution $P(z|\mathbf{m})$.

The weak lensing analysis in Sheldon et al. (2007), on which the results from Johnston et al. (2007) are based, falls somewhere in between these two approaches. While Sheldon et al. (2007) does in fact make use of photometric redshifts, they do not simply assume that the source redshift distribution is identical to the photometric redshift distribution. Rather, they construct a probability distribution $P(z|z_{photo})$, and use this probability to estimate $\left\langle\Sigma_{crit}^{-1}\right\rangle$. As it turns out, evaluating $\left\langle\Sigma_{crit}^{-1}\right\rangle$ in this way leads to results that are nearly identical to those obtained by simply setting $z = z_{photo}$. Thus, even though the approach used in Sheldon et al. (2007) is more sophisticated than the simple case considered in Mandelbaum et al. (2008b), we expect the Sheldon et al. (2007) results to be biased but correctable as prescribed in Mandelbaum et al. (2008b). This correction amounts to a boost of the lensing masses by a factor of 1.18 ± 0.02 ($stat$) ± 0.02 (sys). The statistical error bar in the correction is added in quadrature to the statistical error bar from our fit, which results in

$$\frac{\langle M_{180b}|N\rangle}{10^{14}\ h^{-1}M_\odot} = \exp[0.42 \pm 0.07\ (stat) \pm 0.02\ (sys)] \times (N/20)^{1.18\pm0.09} \tag{A12}$$

These new values for the Johnston et al. (2007) data are in considerably better agreement with those of Mandelbaum et al. (2008a). There remains, however, a systematic 5% difference between the two amplitudes, as well as a small difference $\Delta\alpha_{M|N} = 0.028$ between the two slopes.

A possible culprit for this systematic 5% offset is the difference in how miscentering is accounted for in the data models. The word miscentering refers to the fact that when finding clusters, one will inevitably find clusters that are improperly centered, either due to failures of the cluster finding algorithm, or simply because there is no obvious center of the cluster based on its optical image. Such offsets between the true and assigned centers are problematic because if a cluster is miscentered, the corresponding lensing signal is weakened, resulting in systematically low mass estimates.

To determine whether the remaining offset between Mandelbaum et al. (2008a) and Johnston et al. (2007) is consistent with differences in the miscentering model, we refit our data assuming no errors on the miscentering corrections. We find

$$\langle M_{180b}|N\rangle = \exp(0.42\pm0.04)(N/20)^{1.17\pm0.07}. \tag{A13}$$

with a correlation coefficient $r = -0.15$. Note that these errors are smaller than the errors quoted before, as they should be, given that this new fit does not marginalize over a wide range of miscentering models. By subtracting the two sets of errors in quadrature, we find that the miscentering priors adopted in Johnston et al. (2007) correspond to an error 0.043 in the amplitude and 0.05 in the slope. Thus, the Mandelbaum et al. (2008a) mass measurements are well within the centering error included in the analysis of Johnston et al. (2007).

Nevertheless, it is unclear whether miscentering can in fact account for the difference between the Johnston et al. (2007) and Mandelbaum et al. (2008b) results. More specifically, Mandelbaum et al. (2008a) also performed their analysis including the Johnston et al. (2007) model for miscentering, and find after applying the centering correction their best fit $M_{180b}-N$ relation becomes

$$\langle M_{180b}|N\rangle = \exp(0.53\pm0.07)(N/20)^{1.08\pm0.14}. \tag{A14}$$

Comparing this to equation A12, we find including a miscentering correction in the Mandelbaum et al. (2008a) analysis increases the tension between the two results. Moreover, it suggests that the difference between the two results is due to some other form of systematic difference between the two analysis pipelines. In light of this, we opt for introducing a systematic correction to the Johnston et al. (2007) results of +0.06 and −0.05 for the amplitude and slope respectively. We also introduce systematic errors of the same magnitude as this systematic correction, so that our final result is

$$\langle M_{180b}|N\rangle = [\exp(0.48\pm0.07\,(stat)\pm0.06\,(sys))](N/20)^{1.13\pm0.09\,(stat)\pm0.05\,(sys)}. \tag{A15}$$

Note the central values of the original Johnston et al. (2007) analysis (corrected for photometric redshift bias) as well as the Mandelbaum et al. (2008a) results both with and without miscentering corrections are all encompassed by our systematic error.

Now, in this work we are interested more in the $M_{500c}-N$ (henceforth simply $M-N$) relation than in the $M_{200c}-N$ relation, since it is the former mass which is accessible to X-ray studies. To constrain the $M-N$ relation we use the quoted M_{500c} mass measurements from Johnston et al. (2007), re-scaling their M_{200c} errors to M_{500c} by assuming the relative errors are constant. A fit to the data results in

$$\frac{\langle M|N\rangle}{10^{14}M_\odot} = \exp[0.68\pm0.07](N/40)^{1.11\pm0.08} \tag{A16}$$

with a correlation coefficient $r = 0.45$.

We now boost this expression by factor 1.18 due to the photometric redshift bias correction, and add the systematic corrections +0.06 and −0.05 to the amplitude and slope respectively as per our discussion of the $M_{180b}-N$ relation. We also include a systematic error on the amplitudes and slopes of this same magnitude. We obtain

$$B_{M|N} = 0.91\pm0.07\,(stat)\ \pm0.06\,(sys) \tag{A17}$$
$$\alpha_{M|N} = 1.06\pm0.08\,(stat)\pm0.05\,(sys). \tag{A18}$$

The final systematics we consider here are the purity and completeness of the sample. Now, as long as the completeness is not correlated with mass, completeness should not in any way bias the recovered parameters of the $M-N$ relation, though it obviously affects the error bars due to lower statistics.

The same cannot be said of purity. If only a fraction p of the clusters are actually good matches to real halos in the universe, then a fraction $1-p$ of the clusters will have a lensing signal that is significantly different from the mean signal. As an extreme case, we can consider what happens if a fraction $1-p$ of the clusters had no mass associated with them. In that case, the observed mean mass is simply $M_{obs} = M_{true}/p$ where M_{true} is the true mean, so one should boost the observed masses by a factor of $1/p$ to obtain an unbiased estimate. For $p = 0.9$, this amounts to an increase in $B_{M|N}$ of magnitude $\Delta B_{M|N} = 0.1$. Now, Rozo et al. (2006) showed that the purity of the maxBCG cluster sample is expected to be above 90% over the range or richnesses considered here, and the increase in $B_{M|N}$ quoted above is undoubtedly an overestimate of the necessary correction since even false cluster detections will have excess mass associated with them. In light of this, we have adopted a one-sided systematic error bar $\Delta B_{M|N} = 0.08$ to take into account the impact of purity in the recovered $M-N$ relation. The error bar is one sided since we expect impurities will tend to decrease the observed mean mass. We can, however, turn this prior into a normal double-sided prior by including a systematic correction $\Delta B_{M|N} = 0.04$ to the central value, and setting the systematic error bar to the same magnitude as the central value shift. We can also get a rough estimate for the systematic error on the purity by assuming that the quoted systematic error in the amplitude should be made only in the limit of high or low richness. If that were the case, using the fact the slope is measured over a decade of richness values, the corresponding slope would be

$$\alpha = \frac{1.06\ln(10)+0.08}{\ln(10)} = 1.09 \tag{A19}$$

which amounts to a systematic offset $\Delta\alpha = 0.03$. These systematic error bars are added linearly to our previous systematic error. Our final set of priors for the $M-N$ relation is

$$B_{M|N} = 0.95 \pm 0.07\,(stat) \pm 0.10\,(sys) \tag{A20}$$
$$\alpha_{M|N} = 1.06 \pm 0.08\,(stat) \pm 0.08\,(sys) \tag{A21}$$

with a correlation coefficient $r = 0.45$ between the two statistical errors.

$L_X - N$ Priors

The priors in the $L_X - N$ relation come from repeating the analysis described in Rykoff et al. (2008b), but with L_X defined as the X-ray luminosity in the 0.5-2.0 keV band, and corrected for aperture effects. As in Rykoff et al. (2008b), we restrict this analysis to clusters with $N \geq 30$. We begin by measuring the stacked mean $L_X - N$ relation and scatter on a fixed $1\,h^{-1}$Mpc scale

$$B_{L|N} = 1.69 \pm 0.04\,(stat) \tag{A22}$$
$$\alpha_{L|N} = 1.63 \pm 0.06\,(stat) \tag{A23}$$
$$\sigma_{L|N} = 0.84 \pm 0.03\,(stat) \tag{A24}$$

where we have measured L_X in units of 10^{43} ergs $\times$ s^{-1}, with a pivot point of $N = 40$. We emphasize that the scatter determined above is the total scatter in the observed $L_X - N$ relation that cannot be attributed to Poisson uncertainties in the ROSAT photon counts. In particular, the quoted scatter is affected by possible point source contamination, AGN activity, cool cores, cluster mergers, etc.

There are multiple systematic errors that can affect the derived parameters for the $L_X - N$ relation. These include photometric redshift errors, evolution of the richness parameter N, uncorrelated point sources, cluster mis-centering, and cluster AGN and cool cores. In addition, we need to account for the fraction of cluster flux lost due to our finite aperture and the RASS PSF, in order to compare our results with the luminosity measurements of Vikhlinin et al. (2008). We shall now discuss each of these possible systematic effects.

Rykoff et al. (2008b) find that the accuracy of the maxBCG photo-z estimates is high enough such that any biases are insignificant relative to the statistical uncertainty of the parameter determinations, and can thus be safely ignored. However, Rykoff et al. (2008b) did find significant redshift evolution in the $L_X - N$ relation, well above the expected self-similar evolution. Similar redshift evolution is found in Becker et al. (2007); the reason for the systematic undercounting of cluster members at high redshift is explained in Rozo et al. (2008). We have estimated the effect of this redshift evolution on our derived scatter parameter via a simple Monte Carlo, and confirm that although the apparent evolution is strong, it is insignificant relative to the intrinsic scatter. Therefore, we may also safely ignore this possible systematic effect.

We now take a combined approach to the systematic effects due to cluster mis-centering, a finite aperture, the RASS PSF and uncorrelated point sources. The first three effects are strongly related, in that they all tend to scatter cluster photons out of our initial fixed $1\,h^{-1}$Mpc aperture, and these may affect the normalization, slope, and scatter in the $L_X - N$ relation. Uncorrelated point sources should not affect the mean relation because the large number of stacked sources smooths out the foreground and background. However, when uncorrelated point sources are aligned with individual clusters they may increase the measured scatter by boosting the apparent L_X.

We have estimated the effects of these systematics by running a Monte Carlo with simulated RASS data on top of random backgrounds selected from the area of the RASS photon map that overlaps with the maxBCG mask. We first resample the maxBCG richness function 100 times. Each cluster is given a redshift drawn from the maxBCG redshift distribution, as well as a random postion on the sky selected from the area of the RASS survey that overlaps with the maxBCG mask. After we select the richest 1000 clusters in each realization, each cluster is given a luminosity based on the mean relation from Eqn. A22 and an input intrinsic scatter, $\sigma_{in} = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. Each cluster luminosity is then converted to a number of photon counts according to the RASS exposure at the given point, and scattered by Poisson uncertainties. Then, each cluster is given a position offset according to the maxBCG miscentering distribution described in Johnston et al. (2007, see § 4.3). The cluster profiles are assumed to follow a β model, $S(R) = S_0(1+R^2/R_C^2)^{-3\beta+1/2}$. To ensure we are on similar footing as Vikhlinin et al. (2008), we randomly assign each cluster β model parameters uniformly in the range $0.6 < \beta < 0.7$ and $0.05 < R_C < 0.15\,h^{-1}$Mpc. Finally, the photons are scattered according to the RASS PSF, following the method of Rykoff et al. (2008b, see § 3.3.1). We then calculate the stacked mean relation and scatter as described in Rykoff et al. (2008b).

Figure A1 summarizes the results from our systematic tests. The x-axis shows the input intrinsic scatter, σ_{in}. The y-axis shows the ratio of the input parameter to output parameter for the normalization $B_{L|N}$ (circles), slope $\alpha_{L|N}$ (diamonds), and scatter $\sigma_{L|N}$ (squares). We note that when $\sigma_{in} = 0.0$ then $\sigma_{out} = 0.31 \pm 0.04$, which cannot be displayed on the plot. This is consistent with our expectation that uncorrelated sources may boost the observed scatter, especially with low intrinsic scatter. Overall, we find that (a) the slope $\alpha_{L|N}$ is not significantly biased; (b) at moderate to large scatter ($\sigma_{in} \gtrsim 0.5$) the intrinsic scatter $\sigma_{L|N}$ is not significantly biased; and (c) the output normalization $B_{L|N}$ must be boosted by a factor of 1.20 ± 0.05 to account for the flux lost to miscentering, the finite aperture, and RASS PSF effects. Our priors become then

$$B_{L|N} = 1.87 \pm 0.04\,(stat) \pm 0.05\,(sys) \tag{A25}$$
$$\alpha_{L|N} = 1.63 \pm 0.06\,(stat) \tag{A26}$$
$$\sigma_{L|N} = 0.84 \pm 0.03\,(stat). \tag{A27}$$

In addition to these corrections, we also need to take into account systematic uncertainties due to purity and completeness in the sample. Just as with the weak lensing mass estimates, completeness should not affect the measured $L_X - N$ relation, whereas

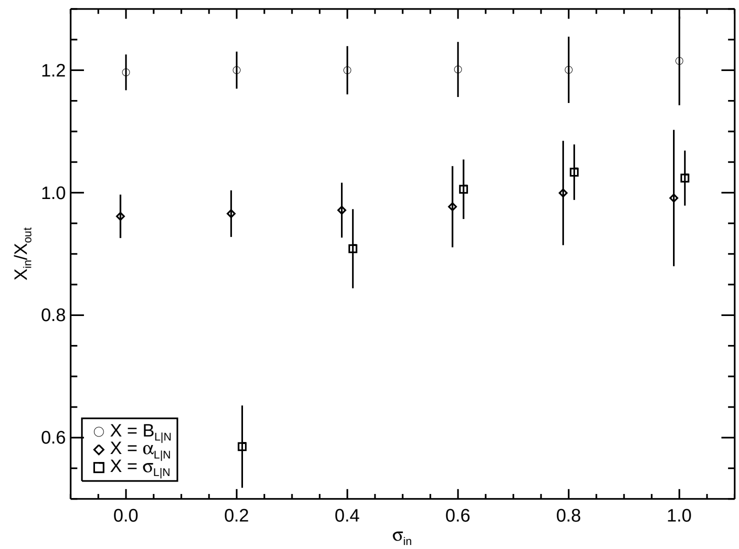

FIG. A1.— Results from systematic error Monte Carlo tests. The x-axis shows the input intrinsic scatter, σ_{in}. The y-axis shows the ratio of the given input parameter to output parameter for the normalization $B_{L|N}$ (circles), slope $\alpha_{L|N}$ (diamonds), and scatter $\sigma_{L|N}$ (squares). We note that when σ_{in} = 0.0 then σ_{out} = 0.31 ± 0.04, which cannot be displayed on the plot.

purity will tend to suppress the X-ray luminosity at fixed richness. Following the same procedure as in appendix A.2, we derive systematic errors $\Delta B_{L|N}$ = 0.04 and $\Delta\alpha_{L|N}$ = 0.05, which we add linearly to our previous systematic error estimates. Finally, we have repeated our scatter analysis using not just the 1000 richest clusters, but also the 2000 richest clusters, in which case we find $\sigma_{L|N}$ = 0.95. To take into account this variation in our analysis, we also introduce a systematic error $\Delta\sigma_{L|N}$ = 0.10. Our final set of priors is

$$B_{L|N} = 1.91 \pm 0.04\,(stat) \pm 0.09\,(sys) \tag{A28}$$

$$\alpha_{L|N} = 1.63 \pm 0.06\,(stat) \pm 0.05\,(sys) \tag{A29}$$

$$\sigma_{L|N} = 0.84 \pm 0.03\,(stat) \pm 0.10\,(sys). \tag{A30}$$

$L_X - M$ Priors

As discussed in section 3, our analysis hinges on the fact that we can use prior knowledge about the $L_X - M$ relation to constrain the $M - N$ relation. Here, we use the results of Vikhlinin et al. (2008) to put priors on the $L_X - M$ relation, which may be summarized as[17]

$$A_{L|M} + 1.361\alpha_{L|M} + 1.5(\sigma^2_{L|M} - 0.40^2) = 2.59 \pm 0.08 \tag{A31}$$

$$\alpha_{L|M} = 1.61 \pm 0.14 \tag{A32}$$

$$\sigma_{L|M} = 0.40 \pm 0.04. \tag{A33}$$

We report a prior on $A_{L|M} + 1.361\alpha_{L|M} + 1.5(\sigma^2_{L|M} - 0.40^2)$ because at $M = 10^{14}\ M_\odot$ the $L_X - M$ parameters derived from the Burenin et al. (2007) sample are correlated. To decouple them, one needs to shift to the statistical pivot point $M = 3.9 \times 10^{14}\ M_\odot$ and introduce the scatter dependence quoted above (Vikhlinin, private communication). These constraints are derived from Chandra observations of clusters in the 400d cluster catalog (Burenin et al. 2007), which allowed Vikhlinin et al. (2008) to measure Y_X and thereby infer cluster mass using the $M - Y_X$ relation. This relation was itself calibrated on a cluster subsample for which masses were derived using the standard hydrostatic equilibrium argument. This last point is very important, since simulations suggest that hydrostatic mass estimates of clusters may be biased low by $\approx 10\% - 30\%$ (see e.g. Evrard 1990; Rasia et al. 2006; Nagai et al. 2007). One way to calibrate such uncertainties is to compare weak lensing mass estimates to hydrostatic mass estimates. There are several examples of this type of approach. For instance, Vikhlinin et al. (2008) have performed such an analysis using the weak lensing mass estimates of Hoekstra (2007), and find $M_{wl} = (1.09 \pm 0.11) M_X$. A similar analysis has been carried out by Mahdavi et al. (2008), who used the weak lensing mass estimates of Hoekstra (2007) and their own analysis of Chandra public data to obtain $M_{wl} = (1.28 \pm 0.15) M_X$. Finally, using XMM X-ray observations and the weak lensing data of Bardeau et al. (2005), Bardeau et al. (2007a), and Dahle (2006), Zhang et al. (2008) find $M_{wl} = (1.21 \pm 0.13) M_X$. Zhang et al. (2008) also note, however, that a histogram of M_{wl}/M_X peaks at a ratio of 1.00 ± 0.05, and that clusters in the tails of the distribution tend to have tight error bars, possibly biasing the error weighted ratio. In light of this, we have opted for a "middle of the road" approach, and introduce a correction factor 1.15 ± 0.15. Our corresponding prior is

$$A_{L|M} + 1.361\alpha_{L|M} + 1.5(\sigma^2_{L|M} - 0.40^2) = 2.45 \pm 0.08\ (stat) \pm 0.23\ (sys) \tag{A34}$$

[17] We have included the appropriate evolution correction for a median redshift z = 0.23, as appropriate for the maxBCG sample.

TABLE B1
MAXBCG MASS FUNCTION DATA

M_{500c}	$\langle dn/d\ln M\rangle$	3.22	3.70	4.26	4.91	5.65	6.50	7.49	8.62	9.92	11.42	13.14	15.12	17.41	20.04	23.07	26.55	30.56
3.22	7.90e-7	0.22	0.82	0.77	0.72	0.66	0.61	0.55	0.50	0.44	0.39	0.35	0.30	0.25	0.21	0.18	0.15	0.12
3.70	5.61e-7	0.82	0.24	0.79	0.76	0.71	0.67	0.62	0.57	0.52	0.46	0.41	0.36	0.31	0.27	0.23	0.19	0.16
4.26	3.92e-7	0.77	0.79	0.27	0.77	0.74	0.70	0.66	0.61	0.57	0.52	0.46	0.41	0.36	0.31	0.26	0.22	0.18
4.91	2.70e-7	0.72	0.76	0.77	0.30	0.75	0.72	0.68	0.64	0.59	0.55	0.50	0.44	0.38	0.34	0.29	0.24	0.20
5.65	1.82e-7	0.66	0.71	0.74	0.75	0.35	0.72	0.69	0.65	0.61	0.57	0.52	0.46	0.41	0.36	0.30	0.26	0.22
6.50	1.21e-7	0.61	0.67	0.70	0.72	0.72	0.41	0.68	0.65	0.62	0.57	0.52	0.47	0.42	0.37	0.32	0.27	0.22
7.49	7.93e-8	0.55	0.62	0.66	0.68	0.69	0.68	0.47	0.64	0.60	0.57	0.52	0.47	0.42	0.37	0.32	0.27	0.23
8.62	5.11e-8	0.50	0.57	0.61	0.64	0.65	0.65	0.64	0.55	0.59	0.55	0.51	0.47	0.41	0.37	0.32	0.27	0.23
9.92	3.24e-8	0.44	0.52	0.57	0.59	0.61	0.62	0.60	0.59	0.65	0.53	0.49	0.45	0.40	0.36	0.31	0.26	0.22
11.42	2.03e-8	0.39	0.46	0.52	0.55	0.57	0.57	0.57	0.55	0.53	0.76	0.47	0.43	0.38	0.34	0.30	0.25	0.21
13.14	1.25e-8	0.35	0.41	0.46	0.50	0.52	0.52	0.52	0.51	0.49	0.47	0.92	0.40	0.36	0.32	0.28	0.24	0.20
15.12	7.61e-9	0.30	0.36	0.41	0.44	0.46	0.47	0.47	0.47	0.45	0.43	0.40	1.11	0.33	0.30	0.26	0.22	0.19
17.41	4.53e-9	0.25	0.31	0.36	0.38	0.41	0.42	0.42	0.41	0.40	0.38	0.36	0.33	1.36	0.26	0.24	0.20	0.17
20.04	2.63e-9	0.21	0.27	0.31	0.34	0.36	0.37	0.37	0.37	0.36	0.34	0.32	0.30	0.26	1.74	0.21	0.19	0.16
23.07	1.48e-9	0.18	0.23	0.26	0.29	0.30	0.32	0.32	0.32	0.31	0.30	0.28	0.26	0.24	0.21	2.22	0.17	0.14
26.55	8.29e-10	0.15	0.19	0.22	0.24	0.26	0.27	0.27	0.27	0.26	0.25	0.24	0.22	0.20	0.19	0.17	2.88	0.12
30.56	4.41e-10	0.12	0.16	0.18	0.20	0.22	0.22	0.23	0.23	0.22	0.21	0.20	0.19	0.17	0.16	0.14	0.12	3.88

Mean and covariance matrix of the maxBCG mass function. Masses are defined using an overdensity of 500 relative to critical, and are measured in units of $10^{14}\ M_\odot$. Space densities are measured in units of Mpc^{-3}. Diagonal terms in the covariance matrix above are set to $\sqrt{C_{i,i}}/\langle n_i\rangle$, and thus represent the fractional uncertainty in the halo space density. Off diagonal terms contain the correlation coefficient $r_{i,j} = C_{i,j}/\sqrt{C_{i,i}C_{j,j}}$ between the various bins. The median redshift of the sample is $z = 0.23$.

$$\alpha_{L|M} = 1.61 \pm 0.14\,(stat) \tag{A35}$$

$$\sigma_{L|M} = 0.40 \pm 0.04\,(stat). \tag{A36}$$

Estimating systematic errors in $\alpha_{L|M}$ and $\sigma_{L|M}$ is difficult. For instance, comparisons with weak lensing masses are not an effective way of assessing systematics because weak lensing mass estimates are so noisy: trying to fit a power law relation between M_{wl} and M_X results in very large errors for the slope of the relation.

One alternative is to consider multiple studies of the $L_X - M$ relation in order to asses how sensitive the recovered parameters are to the analysis pipeline. Unfortunately, such an excercise is far from trivial. One difficulty is the fact that there is very little agreement on the meaning of L_X, with many works focusing on core-excised and/or core-corrected bolometric X-ray lumunisoties (e.g. Bardeau et al. 2007b; Zhang et al. 2007, 2008). Even among those works that also explore the $L_X - M$ relation when L_X is a soft X-ray band luminosity (e.g. Reiprich & Böhringer 2002; Maughan 2007), there are still important differences in the aperture used to estimate L_X. In principle, we could attempt to convert between the various definitions of L_X to try to compare the works against each other, but many of these $L_X - M$ measurements are affected by Malmquist bias, making comparisons to the Vikhlinin et al. (2008) results difficult.

One work that does constrain the the soft X-ray band, non-core excised, Malmquist bias corrected $L_X - M$ relation is Stanek et al. (2006). Unfortunately, the energy band they use is slightly different from that of of Vikhlinin et al. (2008), so even here comparison is not trivial. We expect, however, that at least the scatter and slopes of the $L_X - M$ relation will not be strongly affected by the minor differences between the two L_X definitions. Given our purposes, the interesting thing about the Stanek et al. (2006) results is that they use a very different methodology for constraining the $L_X - M$ relation. In particular, they assume knowledge of cosmological parameters, and then use the observed cluster X-ray luminosity function to constrain $P(L_X|M)$. Assuming their "compromise cosmology", which they argue gives the best results, they find $\alpha_{X|M} = 1.60 \pm 0.05$ and $\sigma_{L|M} = 0.34 \pm 0.10$. These values are in excellent agreement with those of Vikhlinin et al. (2008), and suggest that placing additional systematic errors in the $L_X - M$ parameters is not really necessary at this point.

MASS FUNCTION DATA

Table B1 presents the mean and covariance matrix of the mass function data derived from our analysis. These results represent the state of the art mass function measurements at low redshift from optically derived cluster catalogs. We emphasize we assumed $\Omega_m = 0.27$ and $h = 0.71$, so appropriate rescaling must be applied if the results are to be compared against significantly different cosmologies. Note that the covariance matrix data in table B1 is normalized such that the diagonal entries are the fractional error $\sqrt{C_{i,i}}/\langle n_i\rangle$, while the off diagonal entries are the correlation coefficients $r_{i,j} = C_{i,j}/\sqrt{C_{i,i}C_{j,j}}$. We present the data in this way since it is easier to understand when expressed this way. The actual values for the covariance matrix are easily reconstructed from the data in the table.