THE ASTROPHYSICAL JOURNAL, 520:24–34, 1999 July 20

STOCHASTIC NONLINEAR GALAXY BIASING

AVISHAI DEKEL[1] AND OFER LAHAV[1,2]
Received 1998 June 15; accepted 1999 March 1

ABSTRACT

We propose a general formalism for galaxy biasing and apply it to methods for measuring cosmological parameters, such as regression of light versus mass, the analysis of redshift distortions, measures involving skewness, and the cosmic virial theorem. The common linear and deterministic relation $g = b\delta$ between the density fluctuation fields of galaxies g and mass δ is replaced by the conditional distribution $P(g \mid \delta)$ of these random fields, smoothed at a given scale at a given time. The nonlinearity is characterized by the conditional mean $\langle g \mid \delta \rangle \equiv b(\delta)\delta$, while the local scatter is represented by the conditional variance $\sigma_b^2(\delta)$ and higher moments. The scatter arises from hidden factors affecting galaxy formation and from shot noise unless it has been properly removed. For applications involving second-order local moments, the biasing is defined by three natural parameters: the slope $\hat{b}$ of the regression of g on δ, a nonlinearity $\tilde{b}$, and a scatter σ_b. The ratio of variances $b_{\rm var}^2$ and the correlation coefficient r mix these parameters. The nonlinearity and the scatter lead to underestimates of order $\tilde{b}^2/\hat{b}^2$ and $\sigma_b^2/\hat{b}^2$ in the different estimators of β ($\sim\Omega^{0.6}/\hat{b}$). The nonlinear effects are typically smaller. Local stochasticity affects the redshift-distortion analysis only by limiting the useful range of scales, especially for power spectra. In this range, for linear stochastic biasing, the analysis reduces to Kaiser's formula for $\hat{b}$ (not $b_{\rm var}$), independent of the scatter. The distortion analysis is affected by nonlinear properties of biasing but in a weak way. Estimates of the nontrivial features of the biasing scheme are made based on simulations and toy models, and strategies for measuring them are discussed. They may partly explain the range of estimates for β.

Subject headings: cosmology: theory — dark matter — galaxies: clusters: general — galaxies: distances and redshifts — galaxies: formation — large-scale structure of universe

1. INTRODUCTION

Galaxy "biasing" clearly exists. The fact that galaxies of different types cluster differently (e.g., Dressler 1980; Lahav, Nemiroff, & Piran 1990; Santiago & Strauss 1992; Loveday et al. 1995; Hermit et al. 1996; Guzzo et al. 1997) implies that not all of them are exact tracers of the underlying mass distribution. It is obvious from the emptiness of large voids (e.g., Kirshner et al. 1987) and the spikiness of the galaxy distribution with $\sim 100\ h^{-1}$ Mpc spacing (e.g., Broadhurst et al. 1990), especially at high redshifts (Steidel et al. 1996, 1998), that if the structure has evolved by standard gravitational instability (GI) theory then the galaxy distribution must be biased.

Arguments for different kinds of biasing schemes have been put forward, and physical mechanisms for biasing have been proposed (e.g., Kaiser 1984; Davis et al. 1985; Bardeen et al. 1986; Dekel & Silk 1986; Dekel & Rees 1987; Braun, Dekel, & Shapiro 1988; Babul & White 1991; Lahav & Saslaw 1992). Cosmological simulations of galaxy formation clearly indicate galaxy biasing, even at the level of galactic halos (e.g., Cen & Ostriker 1992; Kauffmann, Nusser, & Steinmetz 1997; Blanton et al. 1998; Somerville et al. 1999). The biasing becomes stronger at higher redshifts (e.g., Bagla 1998a, 1998b; Jing & Suto 1998; Wechsler et al. 1998).

The biasing scheme is interesting by itself as a constraint on the process of galaxy formation, but it is of even greater importance in many attempts to estimate the cosmological density parameter Ω. If one assumes a linear and deterministic biasing relation of the sort $g = b\delta$ between the density fluctuations of galaxies and mass and applies the linear approximation for gravitational instability, $\nabla \cdot \boldsymbol{v} = -f(\Omega)\delta$ with $f(\Omega) \approx \Omega^{0.6}$ (e.g., Peebles 1980), then the observables g and $\nabla \cdot \boldsymbol{v}$ are related via the degenerate combination $\beta \equiv f(\Omega)/b$. Thus, one cannot pretend to have determined Ω by measuring β without a detailed knowledge of the relevant biasing scheme.

It turns out that different methods lead to different estimates of β, sometimes from the same data themselves (for reviews see Dekel 1994, Table 1; Strauss & Willick 1995, Table 3; Dekel, Burstein, & White 1997; Dekel 1999). Most recent estimates for optical and *IRAS* galaxies lie in the range $0.4 \leq \beta \leq 1$.

The methods include, for example, (1) comparisons of local moments of g (from redshift surveys) and δ (from peculiar velocities) or the corresponding power spectra or correlation functions, (2) linear regressions of the fields g and δ or the corresponding velocity fields, (3) analyses of redshift distortions in redshift surveys, and (4) comparisons of the cosmic microwave background dipole with the Local Group velocity as predicted from the galaxy distribution.

In order to sharpen our determination of Ω it is important that we understand the sources for this scatter in the estimates of β. Some of this scatter is due to the different types of galaxies involved, and some may be due to unaccounted for effects of nonlinear gravity and perhaps other sources of systematic errors in the data or the methods. In this paper we investigate the possible contribution to this scatter by nontrivial properties of the biasing scheme—the deviations from linear biasing and the stochastic nature of

[1] Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel; dekel@astro.huji.ac.il.

[2] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, England, UK; lahav@ast.cam.ac.uk.

the biasing scheme. This is done using a simple and natural formalism for general galaxy biasing.

The biasing of density peaks in a Gaussian random field is well formulated (e.g., Kaiser 1984; Bardeen et al. 1986), and it provides a very crude theoretical framework for the origin of galaxy density biasing. In this scheme, the galaxy-galaxy and mass-mass correlation functions are related in the linear regime via

$$\xi_{\rm gg}(r) = b^2 \xi_{\rm mm}(r) , \tag{1}$$

where the biasing parameter b is a constant independent of scale r. However, a much more specific linear biasing model is often assumed in common applications, in which the local density fluctuation fields of galaxies and mass are assumed to be deterministically related via the relation

$$g(\boldsymbol{x}) = b\delta(\boldsymbol{x}) . \tag{2}$$

Note that equation (1) follows from equation (2), but the reverse is not true.

The deterministic linear biasing model is not a viable model. It is based on no theoretical motivation. If $b > 1$, it must break down in deep voids because values of g below -1 are forbidden by definition. Even in the simple case of no evolution in comoving galaxy number density, the linear biasing relation is not preserved during the course of fluctuation growth. Nonlinear biasing, where b varies with δ, is inevitable.

Indeed, the theoretical analysis of the biasing of collapsed halos versus the underlying mass (Mo & White 1996), using the extended Press-Schechter approximation (Bond et al. 1991), predicts that the biasing is nonlinear and provides a useful approximation for its behavior as a function of scale, time, and mass threshold. N-body simulations provide a more accurate description of the nonlinearity of halo biasing (see Fig. 1; Somerville et al. 1999) and show that the model of Mo & White is a good approximation. We provide more details about theoretical, numerical, and observational constraints on the exact shape of nonlinear biasing in § 6, where we estimate the magnitude of nonlinear biasing effects.

It is important to realize that once the biasing is nonlinear at one smoothing scale, the smoothing operation acting on the density fields guarantees that the biasing at any other smoothing scale obeys a different functional form of $b(\delta)$ and is also nondeterministic. Thus, any deviation from the simplified linear biasing model must also involve both scale dependence and scatter.

The focus of this paper is therefore on the consequences of the stochastic properties of the biasing process, which could either be related to the nonlinearity as mentioned above or arise from other sources of scatter. An obvious part of this stochasticity can be attributed to the discrete sampling of the density field by galaxies—the shot noise. In addition, a statistical, physical scatter in the efficiency of galaxy formation as a function of δ is inevitable in any realistic scenario. It is hard to believe that the sole property affecting the efficiency of galaxy formation is the underlying mass density at a certain smoothing scale (larger than the scale of galaxies). For example, the random variations in the density on smaller scales is likely to be reflected in the efficiency of galaxy formation. As another example, the local geometry of the background structure, via the deformation tensor, must play a role too. In this case, the three eigenvalues of the deformation tensor are relevant parameters. Such hidden variables would show up as physical scatter in the density-density relation. A similar scatter is noticeable in the distribution of particular morphological types versus the underlying total galaxy distribution (Lahav & Saslaw 1992). The hidden scatter is clearly seen for halos in simulations including gravity alone (§ 6 below and Fig. 1 based on Somerville et al. 1999) even before the more complex pro-

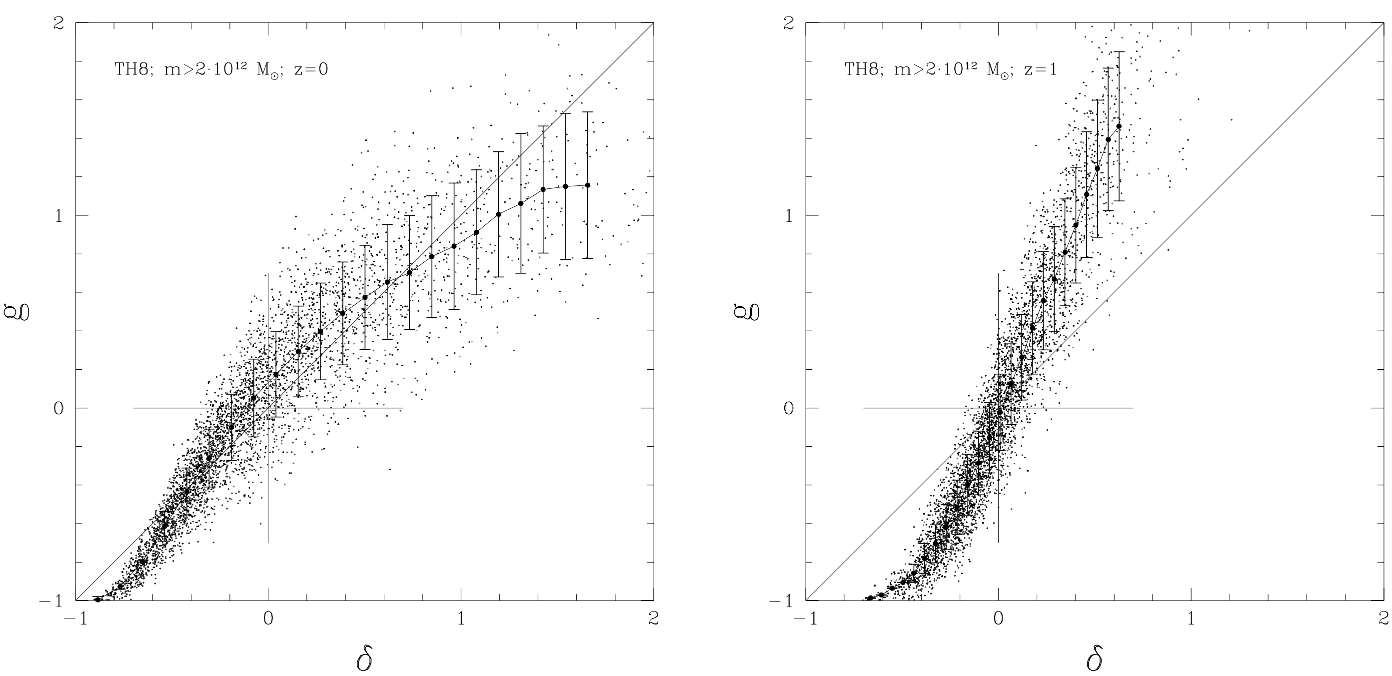

FIG. 1.—Biasing of galactic halos vs. mass in a cosmological N-body simulation, demonstrating nonlinearity and stochasticity. The conditional mean (*solid curve*) and scatter (*error bars*) are marked. The fields smoothed with a top-hat window of radius 8 h^{-1} Mpc are plotted at the points of a uniform grid. *Left-hand panel*: At the time when $\sigma_8 = 0.6$ (e.g., $z = 0$). *Right-hand panel*: At an earlier time when $\sigma_8 = 0.3$ (e.g., $z = 1$). Based on Somerville et al. (1999).

cesses involving gasdynamics, star formation, and feedback affect the biasing and, in particular, its scatter.

In practice, there are two alternative options for dealing with the shot noise component of the scatter. In some cases one can estimate the shot noise and try to remove it prior to the analysis of measuring β. This is sometimes difficult, e.g., because of the finite extent of the galaxies which introduces anticorrelations on small scales. The shot noise is especially large and hard to estimate in the case where the biasing refers to light density rather than number density. The alternative is to treat the shot noise as an intrinsic part of the local stochasticity of the biasing relation without trying to separate it from the physical scatter. The formalism developed below is valid in either case.

In § 2 we present the biasing formalism, separate the effects of nonlinear biasing and stochastic biasing, and apply the formalism to measurements involving local second-order moments of δ and g. In § 3 we derive relations for two-point correlation functions in the presence of local biasing scatter. In § 4 we apply the formalism to the analysis of redshift distortions. In § 5 we address methods involving third-order moments. In § 6 we discuss constraints on the nonlinearity and scatter in the biasing scheme based on simulations, simple models, and observations. In § 7 we summarize our conclusions and discuss our results and future prospects. A very preliminary version of this work has been reported by Dekel (1997).

2. LOCAL MOMENTS

2.1. *Conditional Distribution*

Let $\delta(\boldsymbol{x})$ ($\equiv \delta\rho/\rho$) be the field of mass-density fluctuations and $g(\boldsymbol{x})$ ($\equiv \delta n/n$) the corresponding field of galaxy-density fluctuations (or, alternatively, the field of light-density fluctuations, whose time evolution is less sensitive to galaxy mergers). The fields are both smoothed with a fixed smoothing window, which defines the term "local." The concept of galaxy biasing is meaningful only for smoothing scales larger than the comoving scale of individual galaxies, namely a few Mpc. An example would thus be a top-hat window of radius 8 h^{-1} Mpc, for which the rms fluctuation of optical galaxies is about unity. Our analysis is confined to a specific smoothing scale, at a specific time, and for a specific type of objects.

Assume that both δ and g are random fields, with one-point probability distribution functions (PDFs) $P(\delta)$ and $P(g)$, both of zero mean by definition and of standard deviations $\sigma^2 \equiv \langle \delta^2 \rangle$ and $\sigma_g^2 \equiv \langle g^2 \rangle$. The key idea is to consider the local biasing relation between galaxies and mass to also be a *random* process, specified by the *biasing conditional distribution* $P(g \mid \delta)$ of g at a given δ.

Below, we shall use in several different ways the following lemma relating joint averaging and conditional averaging for any functions $p(g)$ and $q(\delta)$:

$$\langle p(g)q(\delta) \rangle = \langle \langle p(g) \mid \delta \rangle_{g \mid \delta}\, q(\delta) \rangle_\delta \,, \tag{3}$$

where the inner average is over the conditional distribution of g at a given δ and the outer average is over the distribution of δ. This is true because

$$\begin{aligned}\langle p(g)q(\delta) \rangle &= \int\!\!\int dg d\delta P(g,\, \delta) p(g) q(\delta) \\ &= \int d\delta P(\delta) q(\delta) \int dg P(g \mid \delta) p(g) \,,\end{aligned} \tag{4}$$

in which the first equality is by definition and where for the second equality $P(g, \delta)$ has been replaced by $P(g \mid \delta)P(\delta)$ and the double integration has become successive.

2.1.1. *Conditional Mean: Nonlinearity*

Define the *mean biasing function* $b(\delta)$ by the conditional mean,

$$b(\delta)\delta \equiv \langle g \mid \delta \rangle = \int dg P(g \mid \delta) g \,. \tag{5}$$

This function is plotted in Figure 1. This is a natural generalization of the deterministic linear biasing relation, $g = b_1\, \delta$. The function $b(\delta)$ allows for any possible *nonlinear* biasing and fully characterizes it; it reduces to the special case of *linear* biasing when $b(\delta) = b_1$ is a constant independent of δ.

In the following treatment of second-order local moments, we will find it natural to characterize the function $b(\delta)$ by the moments $\hat{b}$ and $\tilde{b}$ defined by

$$\hat{b} \equiv \frac{\langle b(\delta)\delta^2 \rangle}{\sigma^2} \,, \quad \tilde{b}^2 \equiv \frac{\langle b^2(\delta)\delta^2 \rangle}{\sigma^2} \,. \tag{6}$$

In the case of linear biasing they both coincide with b_1. It will be shown that the parameter $\hat{b}$ is the natural generalization of b_1 and that the ratio $\tilde{b}/\hat{b}$ is the relevant measure of nonlinearity in the biasing relation; it is unity for linear biasing and is either larger or smaller than unity for nonlinear biasing.

As can be seen from the definitions of $\hat{b}$ and $\tilde{b}$, this measure is independent of the stochasticity of the biasing. It thus allows one to maintain general nonlinearity while addressing stochasticity.

2.1.2. *Conditional Variance: Stochasticity*

The local *statistical* character of the biasing relation can be expressed by the conditional moments of higher order about the mean at a given δ. Define the *random biasing field* ϵ by

$$\epsilon \equiv g - \langle g \mid \delta \rangle \,, \tag{7}$$

such that its local conditional mean vanishes by definition, $\langle \epsilon \mid \delta \rangle = 0$. The local variance of ϵ at a given δ defines the *biasing scatter function* $\sigma_b(\delta)$ by

$$\sigma_b^2(\delta) \equiv \frac{\langle \epsilon^2 \mid \delta \rangle}{\sigma^2} \,. \tag{8}$$

The scaling by σ^2 is for convenience. The function $\langle \epsilon^2 \mid \delta \rangle^{1/2}$ is marked by error bars in Figure 1. By averaging over δ and using equation (3), one obtains the constant of local *biasing scatter*,

$$\sigma_b^2 \equiv \frac{\langle \epsilon^2 \rangle}{\sigma^2} \,. \tag{9}$$

Thus, to second order, the nonlinear and stochastic biasing relation is characterized locally by three basic parameters: $\hat{b}$, $\tilde{b}$, and σ_b. The parameters $\hat{b}$ and $\tilde{b}/\hat{b}$ refer to the mean biasing and its nonlinearity, while $\sigma_b/\hat{b}$ measures the scatter. This parameterization separates in a natural way the properties of nonlinearity and stochasticity. The formalism simply reduces to the case of linear biasing when $\hat{b} = \tilde{b}$ and to deterministic biasing when $\sigma_b = 0$.

If the biasing conditional distribution, $P(\epsilon \mid \delta)$, is a Gaussian [still allowing $b(\delta)$ and $\sigma_b^2(\delta)$ to vary with δ], then the first- and second-order moments fully characterize the biasing relation. In much of the following we will restrict ourselves to second moments, but we shall see in § 5 that a generalization to higher biasing moments, such as the skewness, is straightforward.

2.2. *Variances and Linear Regression*

From the basic parameters defined above one can derive other useful biasing parameters. A common one is the ratio of *variances*, sometimes referred to as "the" biasing parameter,

$$b_{\rm var}^2 \equiv \frac{\sigma_g^2}{\sigma^2} = \tilde{b}^2 + \sigma_b^2 \ . \qquad (10)$$

The second equality is an interesting result of equation (3). It immediately shows that $b_{\rm var}$ is sensitive both to nonlinearity and to stochasticity and that $b_{\rm var} \geq \tilde{b}$ always. Note the roles of $\tilde{b}^2$ and σ_b^2 as the respective contributions of biasing nonlinearity and biasing scatter to the total scatter in g. This makes $b_{\rm var}$ biased compared with $\hat{b}$,

$$b_{\rm var} = \hat{b}\left(\frac{\tilde{b}^2}{\hat{b}^2} + \frac{\sigma_b^2}{\hat{b}^2}\right)^{1/2} , \qquad (11)$$

by the root of the sum in quadrature of the nonlinearity factor $\tilde{b}/\hat{b}$ and the scatter factor $\sigma_b/\hat{b}$.

Using equation (3), the mean parameter $\hat{b}$ is simply related to the *covariance*,

$$\hat{b}\sigma^2 = \langle g\delta \rangle \ . \qquad (12)$$

Thus, $\hat{b}$ is the slope of the linear regression of g on δ, and it serves as the basic biasing parameter—the natural generalization of the linear biasing parameter b_1. Unlike the variance $\langle g^2 \rangle$ in equation (10), the covariance in equation (12) has no additional contribution from the biasing scatter σ_b.

A complementary parameter to $b_{\rm var}$ is the *linear correlation coefficient*,

$$r \equiv \frac{\langle g\delta \rangle}{\sigma_g \sigma} = \frac{\hat{b}}{b_{\rm var}} = \left(\frac{\tilde{b}^2}{\hat{b}^2} + \frac{\sigma_b^2}{\hat{b}^2}\right)^{-1/2} . \qquad (13)$$

The equalities are based on equation (12) and equation (11), respectively.

The "inverse" regression, of δ on g, yields another biasing parameter:

$$b_{\rm inv} \equiv \frac{\sigma_g^2}{\langle g\delta \rangle} = \frac{b_{\rm var}}{r} = \frac{\hat{b}}{r^2} = \frac{b_{\rm var}^2}{\hat{b}} = \hat{b}\left(\frac{\tilde{b}^2}{\hat{b}^2} + \frac{\sigma_b^2}{\hat{b}^2}\right) . \qquad (14)$$

The parameter $b_{\rm inv}$ is closer to what is measured in reality by two-dimensional linear regression (e.g., Sigad et al. 1998), because the errors in δ are typically larger than in g. Note that $b_{\rm inv}$ is biased compared with $\hat{b}$, with the ratio given by the sum in quadrature of the nonlinearity factor $\tilde{b}/\hat{b}$ and the scatter factor $\sigma_b/\hat{b}$.

It is worthwhile to summarize the relations between the parameters in the two degenerate cases. In the case of *linear* and stochastic biasing, the above parameters reduce to

$$\tilde{b} = \hat{b} = b_1 \ , \quad b_{\rm var} = b_1\left(1 + \frac{\sigma_b^2}{b_1^2}\right)^{1/2} , \quad r = \frac{b_1}{b_{\rm var}} \ ,$$

$$b_{\rm inv} = b_1\left(1 + \frac{\sigma_b^2}{b_1^2}\right) . \qquad (15)$$

Thus, $b_1 \leq b_{\rm var} \leq b_{\rm inv}$.

In the case of nonlinear and *deterministic* biasing they reduce instead to

$$\tilde{b} \neq \hat{b}, \quad \sigma_b = 0 \ , \quad b_{\rm var} = \tilde{b} \ , \quad r = \frac{\hat{b}}{\tilde{b}} \ , \quad b_{\rm inv} = \frac{\tilde{b}^2}{\hat{b}} \ . \qquad (16)$$

Both $b_{\rm inv}$ and $b_{\rm var}$ are biased with respect to $\hat{b}$, and the bias in $b_{\rm inv}$ is always larger. Whether they are biased high or low compared with $\hat{b}$ depends on whether the nonlinearity factor $\tilde{b}/\hat{b}$ is larger or smaller than unity, respectively.

We shall see in § 6 that although $\hat{b}$ and $\tilde{b}$ could significantly differ from unity and from $b(\delta = 0)$, the ratio $\tilde{b}/\hat{b}$, in realistic circumstances, typically obtains values in the range $1.0 \leq \tilde{b}/\hat{b} \leq 1.1$. This means that the effects of nonlinearity are likely to be relatively small.

In the fully degenerate case of linear and deterministic biasing, all the b parameters are the same and only then $r = 1$. Note, again, that the parameters $\hat{b}/\tilde{b}$ and $\sigma_b/\tilde{b}$ nicely separate the properties of nonlinearity and stochasticity, while the parameters $b_{\rm var}$, r, and $b_{\rm inv}$ mix these properties.

In actual applications, the above local biasing parameters are involved when the parameter β is measured from observational data in several different ways. For linear and deterministic biasing this parameter is defined unambiguously as $\beta_1 \equiv f(\Omega)/b_1$. But any deviation from this degenerate model causes us to actually measure different β parameters by the different methods.

For example, it is the parameter $\beta_{\rm var} \equiv f(\Omega)/b_{\rm var}$ which is determined from measurements of σ_g and $\sigma f(\Omega)$. The former is typically determined from a redshift survey, and the latter either from an analysis of peculiar velocity data or from the abundance of rich clusters (with a slightly modified Ω dependence), or by *COBE* normalization of a specific power spectrum shape for mass density fluctuations.

As noted in equation (10), in the case of stochastic biasing $b_{\rm var}$ is always an overestimate of $\tilde{b}$. When the biasing is linear, equation (15), $b_{\rm var}$ is an overestimate of b_1. The corresponding $\beta_{\rm var}$ is thus underestimated accordingly.

A useful way of estimating β (e.g., Dekel et al. 1993; Hudson et al. 1995; Sigad et al. 1998) is via the *linear regression* of the fields in our cosmological neighborhood, e.g., $-\nabla \cdot \boldsymbol{v}(\boldsymbol{x})$ on $g(\boldsymbol{x})$ (or, alternatively, via a regression of the corresponding velocities). In the mildly nonlinear regime, $-\nabla \cdot \boldsymbol{v}(\boldsymbol{x})$ is actually replaced by another function of the peculiar velocity field and its first spatial derivatives, which better approximates the scaled mass density field $f(\Omega)\delta(\boldsymbol{x})$ (e.g., Nusser et al. 1991). The regression that is done, taking the errors on both sides into account, is effectively δ on g, because the errors in $\nabla \cdot \boldsymbol{v}$ (or $f\delta$) are typically more than twice as large as the errors in g (e.g., Sigad et al. 1998). Hence, the parameter that is being measured is close to $\beta_{\rm inv} \equiv f(\Omega)/b_{\rm inv}$. In the case of linear gravitational instability and linear deterministic biasing, the slope of this regression line is simply β_1. Note in equation (15) that in the case of linear and stochastic biasing $b_{\rm inv}$ is an overestimate of b_1. The corresponding β is thus underestimated accordingly in this inverse regression analysis.

3. TWO-POINT CORRELATIONS

For the purpose of redshift-distortion analysis we need to generalize our treatment of stochastic and nonlinear biasing to deal with spatial correlations. Given the random biasing field ϵ, equation (7), we define the two-point biasing-matter cross-correlation function and the biasing autocorrelation function by

$$\xi_{\epsilon m}(r) \equiv \langle \epsilon_1 \delta_2 \rangle , \quad \xi_{\epsilon\epsilon}(r) \equiv \langle \epsilon_1 \epsilon_2 \rangle , \tag{17}$$

where the averaging is over the ensembles at points 1 and 2 separated by r. (Recall that throughout this paper the fields are assumed to be smoothed with a given window.) By the definition of the random biasing field ϵ, and the local scatter, equation (9), at zero lag one has $\xi_{\epsilon m}(0) = 0$ and $\xi_{\epsilon\epsilon}(0) = \sigma_b^2 \sigma^2$. We now define the biasing as *local* if

$$\xi_{\epsilon m}(r) = 0 \quad \text{for any } r , \quad \xi_{\epsilon\epsilon}(r) = 0 \quad \text{for } r > r_b , \tag{18}$$

where r_b is typically on the order of the original smoothing scale. (Some implications of local biasing are discussed by Scherrer & Weinberg 1998.)

A two-point equivalent lemma to equation (3) implies that

$$\langle g_1 \delta_2 \rangle = \langle\langle g_1 \delta_2 \,|\, \delta_1 \delta_2 \rangle_{g|\delta} \rangle_\delta ,$$
$$\langle g_1 g_2 \rangle = \langle\langle g_1 g_2 \,|\, \delta_1 \delta_2 \rangle_{g|\delta} \rangle_\delta . \tag{19}$$

Using these identities, one obtains analogous relations to equations (12) and (10):

$$\xi_{\rm gm}(r) \equiv \langle g_1 \delta_2 \rangle = \langle b(\delta_1)\delta_1 \delta_2 \rangle + \xi_{\epsilon \rm m}(r) , \tag{20}$$

$$\xi_{\rm gg}(r) \equiv \langle g_1 g_2 \rangle = \langle b(\delta_1)\delta_1\, b(\delta_2)\delta_2 \rangle + \xi_{\epsilon\epsilon}(r) . \tag{21}$$

In the case of *linear* and *local* biasing, these become

$$\xi_{\rm gm}(r) = b_1\, \xi_{\rm mm}(r) , \tag{22}$$

$$\xi_{\rm gg}(r) = b_1^2\, \xi_{\rm mm}(r) + \xi_{\epsilon\epsilon}(r), \quad \text{with } \xi_{\epsilon\epsilon}(r) = 0 \quad \text{for } r > r_b . \tag{23}$$

Note that the biasing parameter that appears here is b_1, not to be confused with $b_{\rm var}$ when the biasing is stochastic.

To see how the power spectra are affected by the biasing scatter, we assume, without limiting the generality of the analysis, that the local biasing can be approximated by a step function:

$$\xi_{\epsilon\epsilon}(r) = \begin{cases} \sigma_b^2\, \sigma^2 & r < r_b \\ 0 & r > r_b \end{cases} . \tag{24}$$

Recalling that the power spectra are the Fourier transforms of the corresponding correlation functions,

$$P(k) = 4\pi \int_0^\infty \xi(r)\, \frac{\sin{(kr)}}{kr}\, r^2\, dr , \tag{25}$$

we get for $k \ll r_b^{-1}$, from equations (22) and (23),

$$P_{\rm gm}(k) = b_1\, P_{\rm mm}(k) , \tag{26}$$

$$P_{\rm gg}(k) = b_1^2\, P_{\rm mm}(k) + \sigma_b^2\, \sigma^2 V_b , \tag{27}$$

where V_b is the volume associated with the original smoothing length, r_b. We see that the *local* biasing scatter adds to $P_{\rm gg}(k)$ an *additive constant* at all k-values.

Finally, one can address the effect of the scatter on moments of the fields smoothed at a general smoothing length $r > r_b$. Each of these moments is related to the corresponding power spectrum via an integral of the form

$$\langle \delta\delta \rangle_r = \int d^3k \tilde{W}^2(kr) P(k) , \tag{28}$$

where $\tilde{W}(kr)$ is the Fourier transform of the smoothing window of radius r. Using equation (27) one obtains for linear biasing

$$\sigma_g^2(r) = b_1^2\, \sigma^2(r) + \sigma_b^2\, \sigma^2 (r_b/r)^3 . \tag{29}$$

In the following section we will apply this formalism to the linear analysis of redshift distortions.

For the purpose of an analysis involving *nonlinear* biasing, note that two-point averages as in equations (20) and (21) are calculable once one knows the function $b(\delta)$ and the one- and two-point distributions of the underlying density field δ. It turns out that the relation for $\xi_{\rm gm}$, equation (20), is in fact a simple extension of equation (12) involving only the local moment $\hat{b}$:

$$\xi_{\rm gm}(r) = \hat{b} \xi_{\rm mm}(r) . \tag{30}$$

To prove this, we use $P(\delta_1, \delta_2) = P(\delta_2 \,|\, \delta_1) P(\delta_1)$ to write

$$\langle b(\delta_1)\delta_1\, \delta_2 \rangle = \int d\delta_1\, P(\delta_1) b(\delta_1)\delta_1 \int d\delta_2\, P(\delta_2 \,|\, \delta_1)\delta_2 \tag{31}$$

and use the fact that $\langle \delta_2 \,|\, \delta_1 \rangle = \delta_1\, \xi_{\rm mm}(r)/\sigma^2$. To compute $\hat{b}$ one needs to know only the function $b(\delta)$ and the one-point distribution $P(\delta)$. Higher order moments, like the one in equation (21), would, in general, involve the two-point PDF as well.

4. REDSHIFT DISTORTIONS

A very promising way of estimating β is via redshift distortions in a redshift survey (e.g., Kaiser 1987; Hamilton 1992, 1993, 1995, 1997; Fisher et al. 1994a; Fisher, Scharf, & Lahav 1994; Heavens & Taylor 1995; Cole, Fisher, & Weinberg 1995; Fisher & Nusser 1996; Lahav 1996). Peculiar velocity gradients along the line of sight distort the comoving volume elements in redshift space compared with the corresponding volumes in real space. As a result, a large-scale isotropic distribution of galaxies in real space is observed as an anisotropic distribution in redshift space. The relation between peculiar velocities and mass density depends on Ω, and hence the distortions relative to the galaxy density depend both on Ω and on the galaxy biasing relation. In the deterministic and linear biasing case, this relation involves a single β-parameter (see Hamilton 1997). However, in the general biasing case, the distortion analysis is in principle complicated by the fact that the galaxies play two different roles: they serve both as luminous tracers of the mass distribution as well as test bodies for the peculiar velocity field.

To first order, the local galaxy density fluctuations in redshift space (g_s) and real space (g) are related by $g_s = g - \partial u/\partial r$, where u is the radial component of the galaxy peculiar velocity $\boldsymbol{v}(\boldsymbol{x})$. Assuming no velocity biasing, linear GI theory predicts $\partial u/\partial r = -\mu^2 f(\Omega)\delta$, where μ^2 is a geometrical factor depending on the angle between $\boldsymbol{v}$ and $\boldsymbol{x}$. Thus, the basic linear relation for redshift distortions is

$$g_s = g + f\mu^2 \delta . \tag{32}$$

A general *local* expression for redshift distortions is obtained by taking the mean square

$$\langle g_s g_s \rangle = \langle gg \rangle + 2(f\mu^2)\langle g\delta \rangle + (f\mu^2)^2 \langle \delta\delta \rangle . \tag{33}$$

With our formalism for stochastic biasing, using equations (10) and (12), it becomes

$$\sigma_{g,s}^2 = \sigma_g^2 [1 + 2(f\mu^2) r b_{\rm var}^{-1} + (f\mu^2)^2 b_{\rm var}^{-2}] . \tag{34}$$

This is similar to equation (7) of Pen (1998), in the sense that it involves both $b_{\rm var}$ and r in a nontrivial way and is thus directly affected by the stochasticity of the biasing scheme. However, we shall see that this is true *only* for the *local* moments, at the original smoothing length for which $b_{\rm var}$ and r were defined. When the biasing is stochastic, the situation at nonzero lag is very different.

4.1. *General Linear Redshift Distortions at Nonzero Lag*

The general linear analysis of redshift distortions involves nonlocal analysis. The general expression in terms of correlation functions is obtained straightforwardly from equation (32) by averaging $\langle g_1^s g_2^s \rangle$ over the distributions of δ at a pair of points separated by r:

$$\xi_{\rm gg}^s(r) = \xi_{\rm gg}(r) + 2(f\mu^2)\xi_{\rm gm}(r) + (f\mu^2)^2\xi_{\rm mm}(r) \ . \tag{35}$$

(Recall that our correlation functions and power spectra correspond to the smoothed fields.)

Recalling that the power spectra are the Fourier transforms of the corresponding correlation functions, equation (25), one can equivalently write

$$P_{\rm gg}^s(k) = P_{\rm gg}(k) + 2(f\mu^2)P_{\rm gm}(k) + (f\mu^2)^2 P_{\rm mm}(k) \ . \tag{36}$$

(This expression can alternatively be obtained from the fact that eqs. [33] and [28] are valid for any smoothing scale r.)

Similarly, the spherical harmonic analysis for redshift distortions in linear theory for a flux-limited survey (Fisher et al. 1994b, eq. [11]) can be extended to yield for the mean square harmonics in redshift space:

$$\begin{aligned}\langle |a_{lm}^s|^2 \rangle = \frac{2}{\pi}\int dk k^2 [\,|\Psi_l(k)|^2 P_{\rm gg}(k) \\ + 2f\,|\Psi_l(k)\Psi_l^c(k)|\,P_{\rm gm}(k) \\ + f^2\,|\Psi_l^c(k)|^2 P_{\rm mm}(k)] \ ,\end{aligned} \tag{37}$$

where $\Psi_l(k)$ is a real space window function and $\Psi_l^c(k)$ is a redshift-correction window function, both depending on the selection function and the weighting function. Without elaborating on the details, it is clear that this expression, similar to equation (36), mixes the three different power spectra such that the Ω dependence, in general, may involve more than one unique β.

The crucial question is how to relate the correlation functions, or power spectra, to the biasing scheme. In the case of linear and deterministic biasing, one simply has $P_{\rm gg} = b_1 P_{\rm gm} = b_1^2 P_{\rm mm}$, so the distortion relation reduces to Kaiser's formula,

$$P_{\rm gg}^s = P_{\rm gg}(1 + \mu^2\beta_1)^2 \ , \tag{38}$$

where $\beta_1 \equiv f(\Omega)/b_1$ (and here $b_1 = b_{\rm var}$). In order to obtain more specific distortion relations for the case of stochastic biasing, we shall use the spatial correlations from § 3 in the general distortion relations of the current section.

4.2. *Distortions for Linear, Stochastic and Local Biasing*

At zero lag, by definition, $\xi_{\epsilon\epsilon}(0) = \sigma_b^2\sigma^2$. Then, as in equation (10) for the local moments, $\xi_{\rm gg}(0) = b_{\rm var}^2 \xi_{\rm mm}(0)$ and the general distortion relation for ξ, equation (35), reduces to an equation similar to the local equation (34), in which the second and third terms involve different combinations of r and $b_{\rm var}$ and thus allow us to determine them separately. However, at large separations $r > r_b$, where $\xi_{\epsilon\epsilon}$ vanishes by the assumption of local biasing, one obtains instead, from equations (22) and (23)

$$\xi_{\rm gg}^s(r) = \xi_{\rm gg}(r)[1 + 2(f\mu^2)b_1^{-1} + (f\mu^2)^2 b_1^{-2}] \ . \tag{39}$$

This is the degenerate Kaiser formula, which is very different from the expression for local moments, equation (34). In particular, it is independent of the biasing scatter! It now involves only the mean biasing parameter b_1 (a degenerate combination of r and $b_{\rm var}$), but it contains no information on the stochasticity, σ_b. In terms of b_1, the relation for ξ is identical to the case of deterministic biasing. Thus, the distortion analysis at $r > r_b$ is indeed incapable of evaluating the stochasticity of the process. This is a straightforward result of the assumed locality of the biasing scheme. The biasing scatter at two distant points is uncorrelated and therefore its contribution to $\xi_{\rm gg}$ cancels out.

On the other hand, the redshift distortion analysis is sensitive to the *nonlinear* properties of the biasing relation. A proper analysis would require a nonlinear treatment of the redshift distortions including a nonlinear generalization of the GI relation $\nabla \cdot \boldsymbol{v} = -f\delta$, because the nonlinear effects of biasing and gravity enter at the same order. The result is more complicated than equation (35) but is calculable in principle once one knows the function $b(\delta)$ and the one- and two-point probability distribution functions of δ. The insensitivity to stochasticity remains valid in the case of nonlinear biasing.

Back to the case of linear stochastic biasing. The distortion relation for $P(k)$ becomes more complicated because of the additive term in equation (27). For linear biasing, when substituting equation (27) in equation (36), the terms analogous to the ones involving b_1^{-1} and b_1^{-2} in equation (39) for ξ are multiplied by $[1 - \sigma_b^2\sigma^2 V_b/P_{\rm gg}(k)]$, a function of k. The distortion relation for $P(k)$ is thus affected by the biasing scatter in a complicated way.

However, if the scatter is small, there may be a k range around the peak of $P(k)$ where the additive scatter term in equation (27) is small compared with the rest. In this range the relation reduces to an expression similar to equation (39) for the corresponding power spectra. For example, if $r_b \sim 8\ h^{-1}$ Mpc we have $V_b \sim 2 \times 10^3 (h^{-1}\ {\rm Mpc})^3$, while $P_{\rm mm}(k) \sim 10^4 (h^{-1}\ {\rm Mpc})^3$ at the peak (e.g., Kolatt & Dekel 1997), so a significant k range of this sort is viable, especially if $\sigma_b\sigma \ll b_1$. On the other hand, the scatter term always dominates equation (27) at small and at large k. If $\sigma_b\sigma \sim 1$, then the scatter may dominate already not much below $k \sim 0.01 (h^{-1}\ {\rm Mpc})^{-1}$.

In terms of moments of a general smoothing length $r > r_b$, using equation (29) in equation (33) one obtains a complicated distortion relation again. For small scatter there may be a limited range of scales for which the first term in equation (29) dominates and then the distortions reduce to an equation similar to equation (39) for moments of smoothed fields. At large enough scales, where $P_{\rm mm}$ is rising with k and thus σ^2 is decreasing faster than $\propto x^{-3}$, the scatter becomes dominant.

Note that in order to obtain equation (7) of Pen (1998) from the general linear distortion relation, equation (36), one has to define k-dependent biasing parameters by $P_{\rm gg}(k) = b_{\rm var}(k)^2 P_{\rm mm}(k)$ and $P_{\rm gm}(k) = b_{\rm var}(k) r(k) P_{\rm mm}(k)$. (Note that Pen's β refers to his b_1, which is equivalent to our $b_{\rm var}$, except that he allows it to vary with k). In the case of local biasing, a comparison with our equation (26) and equation (27) yields $b_{\rm var}(k)^2 = b_1^2 + \sigma_b^2\sigma^2 V_b/P_{\rm mm}(k)$ and $b_{\rm var}(k) r(k) = b_1$. In the k range near the peak of $P_{\rm mm}(k)$ where the constant term in equation (27) may be negligible, one has $b_{\rm var}(k) = b_1$ and $r(k) = 1$ and there is indeed no sign of the stochasticity in the distortion relation.

5. SKEWNESS AND THREE-POINT CORRELATIONS

5.1. *Skewness*

We now move to measures of biasing involving third-order moments. Given the biasing random fields ϵ, define the biasing skewness function $S_b(\delta)$ in analogy to $\sigma_b(\delta)$ of equation (8), by

$$S_b(\delta)S \equiv \langle \epsilon^3 \,|\, \delta \rangle \,, \tag{40}$$

where $S \equiv \langle \delta^3 \rangle$. After averaging over δ, the biasing skewness parameter is

$$S_b \equiv \langle \epsilon^3 \rangle / S \,. \tag{41}$$

The biasing parameter that is defined by the ratio of skewness moments is then, based on equation (3) and after some algebra,

$$b_3^3 \equiv \frac{\langle g^3 \rangle}{\langle \delta^3 \rangle} = \frac{\langle \delta^3 b^3(\delta) \rangle}{S} + \frac{3\sigma^2 \langle \delta b(\delta) \sigma_b^2(\delta) \rangle}{S} + S_b \,. \tag{42}$$

In the case of deterministic biasing, $S_b = \sigma_b = 0$, one has

$$b_3^3 = \frac{\langle \delta^3 b^3(\delta) \rangle}{S} \,, \tag{43}$$

which differs from the parameters $\hat{b}$ and $\tilde{b}$ of § 2 because of nonlinear effects.

In the linear case where both $b(\delta)$ and $\sigma_b(\delta)$ are constants, the expression for b_3 reduces to

$$b_3 = b_1 \left(1 + \frac{S_b}{b_1^3} \right)^{1/3} \,. \tag{44}$$

Now, if $S_b = 0$, then $b_3 = b_1$ independently of σ_b. If, on the other hand, $P(g \,|\, \delta)$ is positively skewed, then $b_3 > b_1$.

An interesting quantity involving the skewness and variance of δ is $S_3 \equiv S/\sigma^4$. In the second-order approximation to GI with Gaussian initial fluctuations this quantity is constant in time. For top-hat smoothing and a given power spectrum of an effective power index n at the smoothing scale, this constant is $S_3 = 34/7 - (3 + n)$ (Juszkiewicz, Bouchet, & Colombi 1993). The corresponding quantity involving the moments of g, $S_{3g} \equiv \langle g^3 \rangle / \langle g^2 \rangle^2$, provides an observational measure of biasing (Weinberg 1995):

$$b_{S3}^{-1} \equiv \frac{S_{3g}}{S_3} = \frac{b_3^3}{b_{\rm var}^4} = \frac{\langle \delta^3 b^3(\delta) \rangle / S + 3\sigma^3 \langle \delta b(\delta) \sigma_b^2(\delta) \rangle / S + S_b}{[\langle \delta^2 b^2(\delta) \rangle / \sigma^2 + \sigma_b^2]^2} \,. \tag{45}$$

In the case of deterministic biasing,

$$b_{S3} = \frac{\langle \delta^2 b^2(\delta) \rangle^2 / \sigma^4}{\langle \delta^3 b^3(\delta) \rangle / S} \,. \tag{46}$$

In the case of linear biasing where $b(\delta)$ and $\sigma_b(\delta)$ are constants, this ratio reduces to

$$b_{S3} = b_1 \frac{(1 + \sigma_b^2 / b_1^2)^2}{1 + S_b / b_1^3} \,. \tag{47}$$

The biasing parameter obtained this way thus depends both on σ_b and S_b. If $S_b = 0$, as when $P(g \,|\, \delta)$ is Gaussian, then $b_{S3} = b_1(1 + \sigma_b^2/b_1^2)^2$ and the deviation from b_1 is even larger than that of $b_{\rm inv}$ or $b_{\rm var}$, equation (15). For positive biasing skewness S_b, the parameter b_{S3} may in fact become smaller than b_1.

Szapudi (1998) has shown that, under certain simplifying assumptions, the biasing parameters (taken as two coefficients in a Taylor expansion, eq. [53]) can be determined using three-point statistics (cumulant correlators). The assumptions made are that the biasing is local, deterministic, and scale independent and that redshift distortions are negligible. This approach should be generalized to the more realistic case of stochastic biasing to allow other nontrivial features in the biasing scheme.

5.2. *Cosmic Virial Theorem and Energy Equation*

The cosmic energy equation (CE) (Peebles 1980, § 74; Peebles 1993, eq. [20.11]; Davis, Miller, & White 1997) can be used to determine Ω by relating the observed dispersion of galaxy peculiar velocities to a spatial integral over the galaxy-mass cross-correlation function, $\xi_{\rm gm}(r)$. The observable is the galaxy-galaxy autocorrelation function $\xi_{\rm gg}(r)$, so the corresponding biasing parameter is

$$b_{\rm CE} = \langle gg \rangle / \langle g\delta \rangle \,. \tag{48}$$

At zero lag, $b_{\rm CE} = b_{\rm inv}$ of equation (14). At nonzero lag, $b_{\rm CE}$ can be derived from equations (20) and (21). For linear biasing, equations (22) and (23), one obtains at nonzero lag $b_{\rm CE} = b_1$, which is different from $b_{\rm inv}$ if the biasing is stochastic; see equation (15).

The estimation of Ω via the cosmic virial theorem (CV), as applied to galaxy surveys (Peebles 1980, § 75; Bartlett & Blanchard 1996), relates the observed dispersion of galaxy-galaxy peculiar velocities to a spatial integral over the three-point galaxy-galaxy-mass cross-correlation function, $\xi_{\rm ggm}$ (divided by $\xi_{\rm gg}$). The observable is the three-point galaxy correlation function $\xi_{\rm ggg}$, so the corresponding biasing parameter is

$$b_{\rm CV} = \langle ggg \rangle / \langle gg\delta \rangle \,. \tag{49}$$

At zero lag, using equation (3),

$$b_{\rm CV} = \frac{\langle \delta^3 b^3(\delta) \rangle + 3\sigma^2 \langle \delta b(\delta) \sigma_b^2(\delta) \rangle + S_b S}{\langle \delta^3 b^2(\delta) \rangle + \sigma^2 \langle \delta \sigma_b^2(\delta) \rangle} \,. \tag{50}$$

In the case of deterministic but nonlinear biasing, $b_{\rm CV} = \langle \delta^3 b^3(\delta) \rangle / \langle \delta^3 b^2(\delta) \rangle$, which in general differs from any of the biasing parameters discussed so far. If $b(\delta)$ and $\sigma_b(\delta)$ are constants and the biasing is stochastic, then, at zero lag, $b_{\rm CV} = b_3^3/b_1^2 = b_1(1 + S_b/b_1^3)$.

If the analysis is done on scales smaller than the biasing coherence length r_b, then the local expressions are relevant. Otherwise, one needs to appeal to three-point spatial correlations.

Note that in the case of CV or CE (which are valid on small scales) the measured quantity may be Ω/b or Ω/b^2, depending on the application, rather than β, which is typical in applications based on the linear approximation to gravitational instability.

6. CONSTRAINTS ON THE BIASING RELATION

In the scheme outlined above, the local biasing process at given scale, time, and galaxy type, is characterized by the conditional probability density function $P(g \,|\, \delta)$. The conditional mean, or the function $b(\delta)$, contains the information about the mean biasing (via the parameter $\hat{b}$) and the nonlinear features (e.g., via $\tilde{b}/\hat{b}$). The first additional quantity of interest in the case of nonnegligible scatter in the biasing relation can be the conditional standard deviation, the function $\sigma_b(\delta)$, and its variance over δ, σ_b^2. In order to evaluate the actual effects of nonlinear and stochastic biasing on the various measurements of β, one should first try to constrain these functions or evaluate these parameters from simulations, theoretical approximations, and observations.

6.1. *Preliminary Results from Simulations*

In an ongoing study, Somerville et al. (1999) are investigating the biasing in high-resolution N-body simulations of several cosmological scenarios, both for galactic halos and for galaxies as identified using semianalytic models. Earlier results from simulations were obtained, e.g., by Cen & Ostriker (1992) and in more detail by Mo & White (1996). We refer here to an example of the preliminary results of Somerville et al., in the context of our biasing formalism. As our test case we use a representative cosmological model: $\Omega = 1$ with a τCDM power spectrum which roughly obeys the constraints from large-scale structure. The simulation mass resolution is $2 \times 10^{10}\ M_{\odot}$ inside a box of comoving side 85 h^{-1} Mpc. The present epoch is identified with the time when the rms mass fluctuation in a top-hat sphere of radius 8 h^{-1} Mpc is $\sigma_8 = 0.6$.

Figure 1 is borrowed from Somerville et al. (1999) in order to demonstrate the qualitative features of the biasing scheme. It shows the density fluctuation fields of galactic halos versus mass at the points of a uniform grid at two different times. The halos are selected above a mass threshold of $2 \times 10^{12}\ M_{\odot}$. The fields are smoothed with a top-hat window of radius 8 h^{-1} Mpc. The conditional mean $[\langle g \,|\, \delta \rangle = b(\delta)\delta]$ and the conditional scatter $[\langle \epsilon^2 \,|\, \delta \rangle = \sigma_b^2(\delta)\sigma^2]$ are marked.

The nonlinear behavior in the negative regime, $\delta < 0$, is characteristic of all masses, times, and smoothing scales: the function $\langle g \,|\, \delta \rangle$ is flat near $g = \delta = -1$, and it abruptly steepens toward $\delta = 0$. In the positive regime, $\delta > 0$, the behavior is less robust—it strongly depends on the mass, time, and smoothing scale. The scatter in the figure includes both shot noise and physical scatter which are hard to separate properly. The scatter function $\sigma_b(\delta)$ grows rapidly from zero at $\delta = -1$ to a certain value near $\delta = 0$, and it continues to grow slowly for $\delta > 0$ to an asymptotic value at large δ.

In the case shown at $z = 0$, the nonlinear parameter is $\tilde{b}^2/\hat{b}^2 = 1.08$ and the scatter parameter is $\sigma_b^2/\hat{b}^2 = 0.15$. The effects of stochasticity and nonlinearity in this specific case thus lead to moderate differences in the various measures of β, on the order of 20%–30%. Gasdynamics and other nongravitational processes may extend the range of estimates even further.

A recent hint of the origin of physical scatter in the biasing scheme is provided by Blanton et al. (1998). They find, based on hydrodynamic cosmological simulations, that the local gas temperature, which is an important factor affecting the efficiency of galaxy formation, is not fully correlated with the other dominant factor—the local mass density. They therefore argue that this is a physical hidden variable that contributes to the stochasticity proposed in our current paper. They also detect significant scale dependence in the biasing scheme and identify its main source with the correlation of the temperature with the large-scale gravitational potential.

6.2. *Approximations for Nonlinear Biasing*

Given the distribution $P(\delta)$ of the matter fluctuations, the biasing function $b(\delta)$ should obey by definition at least the following two constraints:

1. $g \geq -1$ everywhere because the galaxy density ρ_g cannot be negative, and $g = -1$ at $\delta = -1$ because there are no galaxies where there is no matter.
2. $\langle g \rangle = 0$ because g describes fluctuations about the mean galaxy density.

An ad hoc example for a simple functional form that automatically obeys the constraint at $\delta = -1$ and reduces to the linear biasing relation near $\delta = 0$ is (e.g., Dekel et al. 1993)

$$\langle g \,|\, \delta \rangle = c(1 + \delta)^b - 1 \ . \tag{51}$$

The constraint $\langle g \rangle = 0$ is yet to be enforced by a specific choice of the factor c as a function of the power b. With $b > 1$, this functional form indeed turns out to provide a reasonable fit to the simulated halo biasing relation in the $\delta < 0$ regime. However, the same value of b does not necessarily fit the biasing relation in the $\delta > 0$ regime, which can require either $b < 1$ or $b > 1$ depending on halo mass, smoothing scale, and redshift.

A better approximation could thus be provided by a combination of two functions like equation (51) with two different biasing parameters $b_{\rm neg}$ and $b_{\rm pos}$ in the regimes $\delta \leq 0$ and $\delta > 0$, respectively. Another useful version of such a combination is

$$\langle g \,|\, \delta \rangle = \begin{cases} (1 + b_0)(1 + \delta)^{b_{\rm neg}} - 1 & \delta \leq 0 \ , \\ b_{\rm pos}\,\delta + b_0 & \delta > 0 \ , \end{cases} \tag{52}$$

which provides an even better fit to the behavior in Figure 1. As mentioned above, the parameter $b_{\rm neg}$ is always larger than unity while $b_{\rm pos}$ ranges from slightly below unity to much above unity. The best fit to Figure 1 at $z = 0$ has $b_{\rm neg} \sim 2$ and $b_{\rm pos} \sim 1$. At high redshift both $b_{\rm neg}$ and $b_{\rm pos}$ become significantly larger.

The nonlinear biasing relation can alternatively be parameterized by a general power series,

$$g = \sum_{n=0}^{\infty} \frac{b_n}{n!}\, \delta^n \ . \tag{53}$$

Since g must average to zero, this power series can be written as

$$\langle g \,|\, \delta \rangle = b_1\, \delta + \tfrac{1}{2}\, b_2(\delta^2 - \sigma^2) + \tfrac{1}{6}\, b_3(\delta^3 - S) + \dots \ , \tag{54}$$

where $\sigma^2 \equiv \langle \delta^2 \rangle$, $S \equiv \langle \delta^3 \rangle$, etc. This determines the constant term b_0. The constraint at -1 provides another relation between the parameters. Therefore, the expansion to third order contains only two free parameters out of four.

In order to evaluate the parameters $\hat{b}$ and $\tilde{b}$ for these nonlinear toy models, we approximate the density PDF as a Gaussian curve, or alternatively as lognormal in $\rho/\bar{\rho} = 1 + \delta$ (e.g., Coles & Jones 1991; Kofman et al. 1994):

$$P(\delta) = \frac{1}{[2\pi \ln(1 + \sigma^2)]^{1/2}} \frac{1}{(1 + \delta)} \times \exp - \frac{[\ln(1 + \delta) - \ln(1 + \sigma^2)^{-1/2}]^2}{2 \ln(1 + \sigma^2)} \ . \tag{55}$$

The only free parameter is σ. The skewness, for example, is $S = 2\sigma^4 + \sigma^6$, etc.

For the nonlinear biasing that is described by Taylor expansion to third order (eq. [54]) and a Gaussian or lognormal density PDF, assuming $b_2 \ll b_1$ and $\sigma \ll 1$, one obtains

$$\frac{\tilde{b}^2}{\hat{b}^2} \simeq 1 + \frac{1}{2}\left(\frac{b_2}{b_1}\right)^2 \sigma^2 \,. \tag{56}$$

This is always larger than unity, but the deviation is small. Alternatively, using the functional form of equation (51), with $b_{\rm neg}$ ranging from 1 to 5, $b_{\rm pos}$ ranging from 0.5 to 3, and a lognormal PDF of $\sigma = 0.7$, we find numerically that $\tilde{b}/\hat{b}$ is in the range 1.0 to 1.15. These two toy models, which approximate the nonlinear biasing behavior seen for halos in the N-body simulations, indicate that despite the obvious nonlinearity, especially in the negative regime, the nonlinear parameter $\tilde{b}/\hat{b}$ is typically only slightly larger than unity. This means that the effects of nonlinear biasing on measurements of β are likely to be relatively small.

6.3. *The Special Case of Gaussian Biasing*

A quick comment on "Gaussian" and "bivariate Gaussian" biasing, which has been used in the recent literature (e.g., Pen 1998). A specific model for nonlinear and stochastic biasing is where the conditional distribution is a Gaussian curve, but allowing $b(\delta)$ and $\sigma_b^2(\delta)$ to vary with δ:

$$P(g \,|\, \delta) \propto \exp - \frac{[g - b(\delta)\delta]^2}{2\sigma^2 \sigma_b^2(\delta)} \,. \tag{57}$$

Based on the N-body simulations, this is a reasonable approximation for galactic halos.

However, the model of *bivariate Gaussian* biasing, which might be tempting because it makes some of the computations easier, is much more restrictive; it is in fact a special case of *linear* biasing. This model assumes that the joint distribution of galaxies and mass is a two-dimensional Gaussian curve,

$$P(\tilde{g}, \tilde{\delta}) \propto \exp - \frac{\tilde{g}^2 - 2r\tilde{g}\tilde{\delta} + \tilde{\delta}^2}{2(1 - r^2)} \,, \tag{58}$$

where $\tilde{g} \equiv g/\sigma_g$ and $\tilde{\delta} \equiv \delta/\sigma$, and with the local biasing correlation coefficient $r =$ const.

If $P(\delta)$ is also a Gaussian curve, $P(\tilde{\delta}) \propto \exp(-\tilde{\delta}^2/2)$, the conditional probability is

$$P(\tilde{g} \,|\, \tilde{\delta}) = \frac{P(\tilde{g}, \tilde{\delta})}{P(\tilde{\delta})} \propto \exp - \frac{(\tilde{g} - r\tilde{\delta})^2}{2(1 - r^2)} \,. \tag{59}$$

This is a one-dimensional Gaussian for $\tilde{g}$, with mean $r\tilde{\delta}$ and variance $(1 - r^2)$. Back to the quantities g and δ, the conditional mean is $\langle g \,|\, \delta \rangle = r b_{\rm var}\, \delta$, where $b_{\rm var} = \sigma_g/\sigma$ as usual. This is thus a special case of *linear* biasing, with a constant linear biasing parameter $b_1 = r b_{\rm var}$ independent of δ, as in equation (15). From the conditional variance of the Gaussian distribution in equation (59), the biasing scatter is also a constant independent of δ, $\sigma_b^2 = b_{\rm var}^2(1 - r^2)$. Based on N-body simulations, linear biasing could be a poor approximation even for galactic halos. The whole biasing scheme is characterized in this case by only two parameters, $b_{\rm var}$ and r, or alternatively b_1 and σ_b, independent of δ.

6.4. *Observations*

Direct constraints on the local biasing field should, in principle, be provided by the data themselves, of galaxy density (e.g., from redshift surveys) versus mass density (e.g., from peculiar velocity surveys, or gravitational lensing). It is a bit early to deduce the nonlinear shape of $b(\delta)$ from these data because of the large errors that they involve at present. However, we note a qualitative example of scatter in the biasing relation in the fact that the smoothed density peaks of the Great Attractor (GA) and Perseus Pisces (PP) are of comparable height in the mass distribution as recovered by POTENT from observed velocities (e.g., Dekel 1994; daCosta et al. 1996; Dekel et al. 1999), but PP is significantly higher than GA in the galaxy map (e.g., Hudson et al. 1995; Sigad et al. 1998). For example, a linear regression of the 12 h^{-1} Mpc smoothed density fields of POTENT mass and optical galaxies in our cosmological neighborhood yields a $\chi^2 \sim 2$ per degree of freedom for the assumed errors in the data (Hudson et al. 1995). One way to obtain a $\chi^2 \sim 1$, as desired, is to assume a biasing scatter of $\sigma_b \sim 0.5$ in the optical density (while $\sigma \sim 0.3$ at that smoothing). With $b_1 \sim 1$, one has $\sigma_b^2/b_1^2 \sim 0.25$. This is only a very crude estimate, and there is yet much to be done along similar lines with future data.

A promising method has been worked out (Sigad & Dekel 1999; see also Dekel 1998) for recovering the mean biasing function $b(\delta)$ and its associated parameters $\hat{b}$ and $\tilde{b}$ from a measured PDF (or counts in cells) of galaxies in a redshift survey. If $g(\delta)$ were deterministic and monotonic, then it could be derived from the cumulative PDFs of galaxies and mass, $C_g(g)$ and $C(\delta)$, via $g(\delta) = C_g^{-1}[C(\delta)]$ (see also Narayanan & Weinberg 1998). It is found for halos in N-body simulations that this is a good approximation for $\langle g \,|\, \delta \rangle$ despite the scatter. The other key point confirmed by a suite of simulations is that $C(\delta)$ is relatively insensitive to the cosmological model or the fluctuation power spectrum and can be approximated for our purpose by a lognormal distribution in $1 + \delta$ (e.g., Bernardeau 1994; Bernardeau & Kofman 1995). Thus, $b(\delta)$ can be evaluated from a measured $C_g(g)$ and the rms σ of mass fluctuations on the same smoothing scale. Since redshift surveys are by far richer than peculiar velocity samples, this method will allow a better handle on $b(\delta)$ than the local comparison of density fields of galaxies and mass. It can be applied to local redshift surveys as well as surveys of objects at high redshift.

Constraints on the biasing scheme can also be obtained by comparing the clustering properties of galaxies of different types in a given redshift surveys (e.g., Lahav & Saslaw 1992). Indeed, partly motivated by the ideas of our current paper, a clear confirmation for nontrivial biasing, nonlinear and/or stochastic beyond shot noise, has recently been reported by Tegmark & Bromley (1998) based on the Las Campanas Redshift Survey.

7. CONCLUSIONS

We have introduced a straightforward formalism for describing the biasing relation between the density fluctuation fields of galaxies and mass, based on the conditional probability function $P(g \,|\, \delta)$. The key feature of this formalism is the natural separation between nonlinear and stochastic effects in the biasing scheme. The nonlinearity is expressed by the conditional mean via the function $b(\delta)$, and the statistical scatter is measured by the conditional standard deviation, $\sigma_b(\delta)$, and higher moments if necessary. For analyses using local moments of second order, the biasing scheme is characterized by three parameters: $\hat{b}$ measuring the mean biasing, $\tilde{b}/\hat{b}$ measuring the effect of nonlinearity, and $\sigma_b/\hat{b}$ measuring the effect of stochasticity.

Deviations from linear and deterministic biasing typically result in biased estimates of the biasing parameter, or the parameter β $(\sim\Omega^{0.6}/b)$, which depend on the actual method of measurement. The nonlinearity and the scatter lead to differences of order $\tilde{b}^2/\hat{b}^2$ and $\sigma_b^2/\hat{b}^2$, respectively, in the different estimators of β using second-order local moments. They typically lead to an underestimate of β with respect to $\hat{\beta} = f(\Omega)/\hat{b}$. Based on N-body simulations and toy models, the effects of nonlinear biasing are typically small, on the order of 20% or less, and the effects of scatter could be somewhat larger. One expects the β parameters from second-order local moments to be biased in the following order: $\beta_{\rm inv} < \beta_{\rm var} < \hat{\beta}$.

The stochasticity affects the redshift-distortion analysis only by limiting the useful range of scales, especially in the analysis involving power spectra. In this range, for *linear* stochastic biasing, the basic *linear* expression reduces to the simple Kaiser formula for $b(\delta) = \hat{b} = b_1$ (not $b_{\rm var}$), and it does not involve the scatter at all. The distortion analysis is in principle sensitive to the nonlinear properties of biasing, but the nonlinear effects, especially at low redshifts, are expected to be weak, and on the same order as the effects of nonlinear gravitational instability. This is good news for the prospects of measuring an unbiased β from redshift distortions in the large redshift surveys of the near future (Two Degree Field and Sloan Digital Sky Survey). A detailed nonlinear analysis of redshift distortions with nonlinear biasing will be reported in a subsequent paper.

More detailed studies of simulations, including different recipes for galaxy formation, are required in order to constrain the parameters of the biasing formalism more accurately. The analysis could also be extended to include nonlocal biasing, using the biasing correlations as defined here.

The study of stochastic and nonlinear biasing should be extended to address the time evolution of biasing because many relevant measurements of galaxy clustering are now being done at high redshifts. As seen in Figure 1, the biasing is clearly a function of cosmological epoch (e.g., M. Rees 1999, private communication; Dekel & Rees 1987; Mo & White 1996; Steidel et al. 1996, 1998; Bagla 1998a, 1998b; Matarrese et al. 1997; Wechsler et al. 1998; Peacock 1998). In particular, if galaxy formation is limited to a given epoch and the biasing is linear, one can show (e.g., Fry 1996) that the linear biasing factor b_1 would eventually approach unity as a simple result of the continuity equation. Tegmark & Peebles (1998) have generalized the analytic study of time evolution to the case of stochastic but still linear biasing and showed how $b_{\rm var}$ and r approach unity in this case. Analytic attempts to study the evolution of mildly nonlinear stochastic biasing have been reported recently (Taruya, Koyama, & Soda 1998; Taruya & Soda 1998; Catelan et al. 1998; Catelan, Matarrese, & Porciani 1998; Sheth & Lemson 1998). These studies can be extended to the general nonlinear case using our formalism. The simulations of Somerville et al. (1998) are aimed at this goal.

More accurate measurements of peculiar velocities in our greater cosmological neighborhood, and careful comparisons with the galaxy distribution, promise to allow improved observational estimates of the biasing scatter in the future. The reconstruction of the large-scale mass distribution based on weak gravitational lensing (Van Waerbeke 1998, 1999; Schneider 1998; Kaiser et al. 1998) is also becoming promising for this purpose.

The main moral of this paper is that in order to put any measurement of β in cosmological perspective, and in particular when trying to use it for an accurate measurement of the cosmological parameter Ω, one should consider the effects of nonlinear and stochastic biasing and the associated complications of scale dependence, time dependence, and type dependence. The current different estimates are expected to span a range of $\sim 30\%$ in β because of stochastic and nonlinear biasing. The analysis of redshift distortions seems to be most promising; once it is limited to the appropriate range of scales, the analysis is independent of stochasticity and the nonlinear effects are expected to be relatively small.

We thank Adi Nusser, Ue-Li Pen, Martin Rees, Rachel Somerville, Michael Strauss, Max Tegmark, and S. D. M. White for stimulating discussions. This research was supported in part by the US-Israel Binational Science Foundation grant 95-00330 and by the Israel Science Foundation grants 950/95 and 546/98.

REFERENCES

Babul, A., & White, S. D. M. 1991, MNRAS, 253, L31
Bagla, J. S. 1998a, MNRAS, 297, 251
———. 1998b, MNRAS, 299, 417
Bardeen, J., Bond, J. R., Kaiser, N., & Szalay, A. 1986, ApJ, 304, 15
Bartlett, J. B., & Blanchard, A. 1996, A&A, 464, 805
Bernardeau, F. 1994, A&A, 291, 697
Bernardeau, F., & Kofman, L. 1995, ApJ, 443, 479
Blanton, M., Cen, R., Ostriker, J. P., & Strauss, M. A. 1998, ApJ, submitted (astro-ph/9807029)
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440
Braun, E., Dekel, A., & Shapiro, P. 1988, ApJ, 328, 34
Broadhurst, T. J., Ellis, R. S., Koo, D. C., & Szalay, A. S. 1990, Nature, 343, 726
Catelan, P., Lucchin, F., Matarrese, S., & Porciani, C. 1998, MNRAS, 297, 692
Catelan, P., Matarrese, S., & Porciani, C. 1998, ApJ, 502, L1
Cen, R., & Ostriker, J. P. 1992, ApJ, 399, L113
Cole, S., Fisher, K. B., & Weinberg, D. 1995, MNRAS, 275, 515
Coles, P., & Jones, B. 1991, MNRAS, 248, 1
da Costa, L. N., Freudling, W., Wegner, G., Giovanelli, R., Haynes, M. P., & Salzer, J. J. 1996, ApJ, 468, L5
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
Davis, M., Miller, A., & White, S. D. M. 1997, ApJ, 460, 63
Dekel, A. 1994, ARA&A, 32, 371
———. 1997, in Proc. 3d ESO-VLT Workshop, Galaxy Scaling Relations: Origins, Evolution and Applications, ed. L. da Costa & A. Renzini (Berlin: Springer), 245
———. 1998, in IAP Symp. 14, Large-Scale Surveys, ed. S. Colombi, Y. Mellier, & B. Raban (Gif-sur-Yvette: Editions Frontières), 47
Dekel, A. 1999, in Formation of Structure in the Universe, ed. A. Dekel & J. P. Ostriker (Cambridge: Cambridge Univ. Press), in press
Dekel, A., Bertschinger, E., Yahil, A., Strauss, M. A., Davis, M., & Huchra J. P. 1993, ApJ, 412, 1
Dekel, A., Burstein, D., & White, S. D. M. 1997, in Critical Dialogues in Cosmology, ed. N. Turok (Singapore: World Scientific), 175
Dekel, A., Eldar, A., Kolatt, T., Yahil, A., Willick, J. A., Faber, S. M., Courteau, S., & Burstein, D. 1999, ApJ, in press
Dekel, A., & Rees, M. J. 1987, Nature, 326, 455
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Dressler, A. 1980, ApJ, 236, 351
Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994a, MNRAS, 267, 927
Fisher, K. B., & Nusser, A. 1996, MNRAS, 279, L1
Fisher, K. B., Scharf, C. A., & Lahav, O. 1994b, MNRAS, 266, 219
Fry, J. N. 1996, ApJ, 461, L65
Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37
Hamilton, A. J. S. 1992, ApJ, 385, L5
———. 1993, ApJ, 406, L47
———. 1995, in Clustering in the Universe, Proc. 30th Rencontres de Moriond, ed. S. Maurogordato, C. Balkowski, C. Tao, & J. Trân Thanh Vân (Gif-sur-Yvette: Editions Frontières), 143
———. 1997, MNRAS, 289, 285
Heavens, A. F., & Taylor, A. N. 1995, MNRAS, 275, 483
Hermit, S., Santiago, B. X., Lahav, O., Strauss, M. A., Davis, M., Dressler, A., & Huchra, J. P. 1996, MNRAS, 283, 709
Hudson, M. J., Dekel, A., Courteau, S., Faber, S. M., & Willick, J. A. 1995, MNRAS, 274, 305
Jing, Y. P., & Suto, Y. 1998, ApJ, 494, L5

Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, ApJ, 412, L9
Kaiser, N. 1984, ApJ, 284, L9
———. 1987, MNRAS, 227, 1
Kaiser, N., Wilson, G., Luppino, G., Kofman, L., Gioia, I., Metzger, M., & Dahle, H. 1998, ApJ, submitted (astro-ph/9809268)
Kauffman, G., Nusser, A., & Steinmetz, M. 1997, MNRAS, 286, 795
Kirshner, R. P., Oemler, A. Jr., Schechter, P. L., & Shectman, S. A. 1987, ApJ, 314, 493
Kofman, L., Bertschinger, E., Gelb, J., Nusser, A., & Dekel, A. 1994, ApJ, 420, 44
Kolatt, T., & Dekel, A. 1997, ApJ, 479, 592
Lahav, O. 1996, Helvetica Physica Acta, 96, 388
Lahav, O., Nemiroff, R. J., & Piran, T. 1990, ApJ, 350, 119
Lahav, O., & Saslaw, W. 1992, ApJ, 396, 430
Loveday, J., et al. 1995, ApJ, 442, 457
Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, MNRAS, 286, 115
Mo, H., & White, S. D. M. 1996, MNRAS, 282, 347
Narayanan, V. K., & Weinberg, D. H. 1998, ApJ, 508, 440
Nusser, A., Dekel, A., Bertschinger, E., & Blumenthal, G. R. 1991, ApJ, 379, 6
Peacock, J. A. 1998, Philos. Trans. R. Soc. London A, submitted (astro-ph/9805208)
Peebles, P. J. E. 1980, Large-Scale Structure in the Universe (Princeton: Princeton Univ. Press)
———. 1993, Principles of Physical Cosmology (Princeton: Princeton Univ. Press)
Pen, U.-L. 1998, ApJ, 504, 601
Santiago, B. X., & Strauss, M. A. 1992, ApJ, 387, 9
Scherrer, R. J., & Weinberg, D. H. 1998, ApJ, 504, 607
Schneider, P. 1998, ApJ, 498, 43
Sheth, R. K., & Lemson, G. 1998, MNRAS, in press
Sigad, Y., & Dekel, A. 1998, in preparation
Sigad, Y., Eldar, A., Dekel, A., Strauss, M. A., & Yahil, A. 1998, ApJ, 495, 516
Somerville, R., Sigad, Y., Lemson, G., Dekel, A., Colberg, J., Kauffmann, G., & White, S. D. M. 1999, in preparation
Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., Pettini, M., & Kellogg, M. 1998, ApJ, 492, 428
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
Strauss, M. A., & Willick, J. A. 1995, Phys. Rep., 261, 271
Szapudi, I. 1998, MNRAS, 300, L35
Taruya, A., Koyama, K., & Soda, J. 1998, ApJ, 510, 541
Taruya, A., & Soda, J. 1998, ApJ, submitted (astro-ph/9809204)
Tegmark, M., & Bromley, B. C. 1998, ApJ Lett., submitted (astro-ph/9809324)
Tegmark, M., & Peebles, P. J. E. 1998, ApJ, 500, L79
Van Waerbeke, L. 1998, A&A, 334, 1
———. 1999, in IAP Symp. 14, Large-Scale Surveys, ed. S. Colombi, Y. Mellier, & B. Raban (Gif-sur-Yvette: Editions Frontières), 189
Wechsler, R. H., Gross, M. A. K., Primack, J. R., Blumenthal, G. R., & Dekel, A. 1998, ApJ, 506, 19
Weinberg, D. H. 1995, in Wide-Field Spectroscopy and the Distant Universe, ed. S. J. Maddox & A. Aragon-Salamanaca (Singapore: World Scientific), 129