Statistics of the Dark Matter Halo Distribution in Cosmic Density Fields

vorgelegt von Rigoberto Angel Casas Miranda

aus Bogotá, Kolumbien

München, den 4. Februar 2002

1. Gutachter: Prof. Dr. Gerhard Börner

2. Gutachter: Prof. Dr. Axel Schenzle

Tag der mündlichen Prüfung: 11. Juli 2002

Für Astrid

Contents

Zusammenfassung

Gemäß den aktuellen kosmologischen Modellen besteht der Großteil der Masse im Universum aus dunkler Materie. Aus früheren Studien ist es bekannt, daß Galaxien verschiedener Typen auf verschiedene Weise verteilt sind. Die räumliche Verteilung von Galaxien folgt nicht der Verteilung der Masse. Die Relation zwischen der Galaxienverteilung und der Massenverteilung wird „Bias der Galaxienverteilung´´ genannt.

Laut den derzeitigen Modellen für die Bildung von Galaxien entstehen die Galaxien durch das Abkühlen und die Kondensation des baryonischen Gases innerhalb der Potentialtöpfe von virialisierten Klumpen aus dunkler Materie (dunkle Halos). Obwohl die hydrodynamischen Prozesse, die an der Entstehung von Galaxien beteiligt sind, noch wenig verstanden sind, wird angenommen, daß diese Prozesse für die Entstehung einzelner Objekte relevant sind und daß sie möglicherweiser keine bedeutende Rolle bei der gesamten räumlichen Verteilung der Galaxien spielen. Das bedeutet, daß das Problem der Verteilung der Galaxien und des sogenannten Bias der Galaxienverteilung gut untersucht werden kann, indem man die Verteilung von dunklen Halos betrachtet. Diese Annäherung ist sehr praktisch, weil bei der Haloentstehung und der Verteilung die Gravitation der einzige beteiligte physikalische Prozeß ist.

In dieser Arbeit beschäftige ich mich mit den Eigenschaften der räumlichen Verteilung von dunklen Halos auf kosmischen Dichtefeldern. Die Analyse wird in zwei Hauptteilen durchgeführt. Im ersten Schwerpunkt studiere ich deterministische Bias- Modelle, die auf einem sphärischen Kollapsmodell, sowie auf einem ellipsoidförmigen Kollapsmodell beruhen. Im zweiten Teil meiner Arbeit konzentriere ich mich auf die stochastische Beschaffenheit des Bias der Halo- und Galaxien-Verteilung unter Verwendung der bedingten Wahrscheinlichkeitsfunktion.

Ich studiere den deterministischen Bias der Haloverteilung mit Hilfe von verschiedenen Modellen für die Bias Relation zwischen dunklen Halos und der darunterliegenden Ma-

terie. Mit der Benutzung von N-Körper Simulationen mit hoher Auflösung prüfe ich einige theoretische Modelle für die Streuung und für höherwertige Momente der Halo Verteilung in Modellen mit kalter dunkler Materie (CDM, englisch *cold dark matter*). Ich habe herausgefunden, daß die theoretischen Modelle des Bias, die auf einem sphärischen Kollapsmodell beruhen, die simulierten *counts-in-cells* Momente für Halos mit Massen größer als $M^\star$ ziemlich genau beschreiben. $M^\star$ wird als die Massenskala, auf der die Fluktuation des Dichtefeldes ein *rms* von ungefähr 1 hat, definiert. Eine bedeutende Verbesserung der theoretischen Beschreibung der simulierten *counts-in-cells* Momente für unter-M^* Halos wird erzielt, wenn ein ellipsoidförmiges Kollapsmodell anstelle eines sphärischen für die Definition von dunklen Halos benutzt wird. Beide Versionen der Modelle sind besonders genau in der Beschreibung der *counts-in-cells* Momente der Nachkommen von Halos, die bei hohen Rotverschiebungen ausgewählt worden sind. Deswegen sind diese Bias-Modelle ziemlich nützlich für die Interpretierung der Momente der Galaxienverteilung.

Als eine Anwendung der Bias-Modelle berechne ich die Voraussage der Modelle für die höherwertigen Momente der Verteilung der *Lyman break Galaxien* und deren Nachkommen. Es wird angenommen, daß die *Lyman break Galaxien* im Zentrum der massivsten Halos bei der Rotverschiebung $z \sim 3$ entstehen. Ich habe festgestellt, daß, obwohl der lineare Bias-Parameter b stark von der angenommenen Kosmologie abhängt, die Werte der höherwertigen Momente praktisch dieselben in beiden ΛCDM und τCDM Modelle sind. Folglich können die höherwertigen Momente der räumlichen Verteilung dieser Objekte die kosmologische Parameter nicht eingrenzen.

Außerdem betrachte ich die stochastische Natur der Bias Relation vom Gesichtspunkt der bedingten Wahrscheinlichkeitsfunktion aus. Die stochastische Natur der Verteilung von dunklen Halos in einem kosmischen Dichtefeld zeigt sich in der Verteilungsfunktion $P_V(N|\delta_m)$, die die Wahrscheinlichkeit angibt, N Halos in einem Volumen V mit Massendichtekonstrast δ_m zu finden. Diese bedingte Wahrscheinlichkeitsfunktion spezifiziert vollständig die Bias-Relation in einem statistischen Sinn.

Die Annahme, daß die Population von Galaxien und dunklen Halos durch einen Poisson-prozeß (d.h. die bedingte Wahrscheinlichkeit Funktion hat die Form einer Poissonverteilung) erzeugt wurde, hat keine physikalische Unterstützung. Deshalb ist es wichtig zu prüfen, ob andere Verteilungsfunktionen die bedingte Wahrscheinlichkeit besser beschreiben können. Ich benutze drei Funktionen, zusammen mit der Poissonfunktion,

um es nachzuprüfen, wie sie die bedingte Wahrscheinlichkeit aus N-Körper Simulationen hoher Auflösung reproduzieren. Diese drei Funktionen sind die Gauss, die Lognormal und die Thermodynamische Verteilung. Die Thermodynamische Verteilung wurde in den achtziger Jahren aus thermodynamischen Argumenten entwickelt.

Ich fand, daß die bedingten Wahrscheinlichkeitsfunktionen für Halo Massen von einer Gaussfunktion besser beschrieben werden, und daß $P_V(N|\delta_m)$ significant nicht-Poisson ist. Das Verhältnis zwischen der Streuung und dem Erwartungswert geht von ~ 1 (Poisson) bei $1 + \delta_m \ll 1$ bis < 1 (unter-Poisson) bei $1 + \delta_m \sim 1$ bis > 1 (über-Poisson) für $1 + \delta_m \gg 1$. Es stellte sich heraus, daß der Mittelwert der Biasrelation durch Halo Bias Modelle, die auf dem Press-Schechter Formalismus beruhen, gut beschrieben wird. Die unter-Poisson Streuung kann als eine Folge von Halo-Ausschließung begründet werden, während die über-Poisson Streuung bei hohen δ_m Werte durch Halo-Bündelung begründet werden kann. Ein einfaches phänomenologisches Modell für die Streuung der Bias-Relation, als Funktion von δ_m, wird vorgeschlagen. Galaxienkataloge, die mit Hilfe semi-analytischer Modelle aus der N-Körper Simulationen erzeugt worden sind, wurden benutzt, um das Verhalten des Bias der Galaxienverteilung zu untersuchen. Der Bias der Galaxienverteilung, die aus semi-analytischen Modellen der Galaxienentstehung abgeleitet wird, zeigt ein ähnliches stochastisches Verhalten wie der von dunklen Halos. Die bedingte Wahrscheinlichkeit für Galaxien wird durch eine Gaussfunktion gut beschrieben.

Diese Resultate haben wichtige Implikationen bei den Deutungen der Verteilung von Galaxien in Bezug auf das zugrundeliegenden Dichtefeld. Um die Eigenschaften der Massenverteilung im Universum aus statistischen Maßen der Galaxienverteilung abzuleiten, ist es notwendig, zuerst die stochastische Natur des Bias der Galaxienverteilung zu verstehen.

Die Hauptteile dieser Arbeit befinden sich in den Artikeln Casas-Miranda et al. (2002) und Casas-Miranda et al. (2002 in Vorbereitung).

Summary

In the current picture of the universe it is assumed that the mass content in the universe is dominated by a dark matter component. From earlier studies it is known that galaxies of different type cluster in different ways, which means that they do not trace the underlying mass exactly. The relationship between the galaxy and mass distributions is known as "Galaxy bias".

In the current framework of galaxy formation it is assumed that galaxies are formed by the cooling and condensation of the baryonic gas trapped within the potential wells of virialized dark matter clumps (dark matter haloes). Although the hydrodynamical processes involved in the formation of galaxies are still poorly understood, it is assumed that these processes are mainly relevant for the formation of individual objects and that they do not play a significant role in the overall spatial distribution of galaxies. That means that the problem of galaxy clustering and galaxy biasing can be fairly approximated by investigating the clustering of dark matter haloes. This approach is very convenient because the physics involved in the process of halo formation and clustering are quite simple: there is only gravity involved.

In this thesis I study the clustering properties of the spatial distribution of dark matter haloes in cosmic density fields. The analysis is carried out in two main parts. The first part corresponds to the study of deterministic halo-mass bias models, based on a spherical collapse model as well as on an ellipsoidal collapse model. The second part corresponds to the study of the stochasticity in the halo-mass bias relation from the point of view of the conditional probability.

I study the deterministic halo-mass bias relation using several deterministic models for the bias relation between dark matter haloes and the underlying mass. Using high-resolution N-body simulations, I test some theoretical models for the variance and higher-order moments of the dark halo distribution in Cold Dark Matter universes. I have

found that the theoretical biasing models based on the spherical collapse describe quite accurately the simulated counts-in-cells moments for haloes with masses larger than $M^\star$ (the mass scale on which the fluctuation of the density field has a rms about 1). Significant improvement can be achieved for sub-M^* haloes if an ellipsoidal collapse model is used instead of the spherical model in defining dark haloes. Both versions of the models (i.e. based either on spherical collapse or on ellipsoidal collapse) are particularly accurate for the descendants of haloes selected at high redshift, and so are quite useful in interpreting the higher-order moments of galaxies. As an application I use the theoretical model to predict the higher-order moments, at a fixed scale, of the Lyman break galaxies, assumed to form in the center of the most massive haloes at redshift ~ 3, observed at $z \approx 3$ and their descendants at lower redshifts. I have found that, although the linear bias parameter b depends strongly on the cosmology adopted, the values of the higher-order moments are practically the same in both ΛCDM and τCDM dark matter universes and therefore the higher-order moments from the spatial distribution of these objects cannot constrain cosmological parameters.

In addition, I investigate the stochastic nature of the halo-bias relation from the point of view of the conditional probability. The stochasticity in the distribution of dark haloes in the cosmic density field is reflected in the distribution function $P_V(N|\delta_m)$ which gives the probability of finding N haloes in a volume V with mass density contrast δ_m. This conditional probability completely specifies the bias relation in a statistical sense.

It has been widely accepted that the population of galaxies (and dark matter haloes) is obtained from the underlying mass distribution as a Poisson process (i.e. the conditional probability has the form of a Poisson distribution). This assumption has no physical support and, therefore, it is important to test whether other functions can describe better the conditional probability. I use three distribution functions, along with the Poisson one, to investigate how they reproduce the conditional probability obtained from high resolution N-body simulations. These three functions correspond to the Gaussian, the Lognormal and the Thermodynamic distributions. The last one was developed in the 80's based on thermodynamic arguments.

It has been found that the halo-mass conditional probability functions are best described by a Gaussian function and that $P_V(N|\delta_m)$ is significantly non-Poisson. The ratio between the variance and the mean goes from ~ 1 (Poisson) at $1 + \delta_m \ll 1$ to < 1 (sub-

Poisson) at $1 + \delta_m \sim 1$ to > 1 (super-Poisson) at $1 + \delta_m \gg 1$. The mean bias relation has been found to be well described by halo bias models based on the Press-Schechter formalism. The sub-Poisson variance can be explained as a result of halo-exclusion while the super-Poisson variance at high δ_m may be explained as a result of halo clustering. A simple phenomenological model is proposed to describe the behavior of the variance as a function of δ_m. Galaxy catalogues obtained from the simulations using semi-analytical models of galaxy formation were used to investigate the behavior of the galaxy-mass bias relation. It has been found that the galaxy distribution in the cosmic density field predicted by semi-analytic models of galaxy formation shows similar stochastic behavior as dark matter haloes do. It has been found also that the conditional probability for galaxies is well described by a Gaussian function.

These results have important implications in the interpretations of galaxy clustering in terms of the underlying density field, as discussed in chapter 2. Thus, in order to infer the properties of the mass distribution in the Universe from statistical measures of the galaxy distribution, it is necessary to understand first the stochastic nature of galaxy biasing.

The main results of this thesis are under the process of publication in the articles Casas-Miranda et al. (2002) and Casas-Miranda et al. (2002, in preparation).

Chapter 1

Introduction

1.1 Motivation

One of the most important questions in modern cosmology is to understand how the structure, that is the distribution of galaxies, galaxy clusters and of the underlying matter, in the universe has been formed. In order to address this question it is necessary to have well defined quantities to describe the structure present in the universe, as well as self consistent theoretical models of structure formation. Normally theoretical models to describe large-scale structure in the universe are based on some random or stochastic initial conditions. Thus, it is required to interpret clustering data, whether from observations or from numerical simulations, in a statistical way.

The distribution of galaxies is the directly observable part of the structure in the universe. Thus, observationally, the large scale structure of the universe can be investigated by analyzing the statistics of the distribution of galaxies. In the last years, there has been an increasing progress in the accuracy and sky coverage of galaxy surveys. As the quality and quantity of the data increases, it is necessary to improve the theoretical models of structure formation, as well as the statistics used to quantify the clustering. There are several statistical measures of galaxies, each of them having a particular utility depending on the special features of the structure one aims to analyze. Among the most common quantities used to measure local galaxy clustering we find the N-point correlation functions as well as the distribution of counts-in-cells and its moments.

In order to construct a good theoretical model of structure formation one should be able to describe all the physical processes involved in the formation and clustering of galaxies. In addition the model must be able to reproduce the clustering patterns observed from galaxy catalogues. Here the most challenging problem is to understand the relationship between the galaxy distribution and the underlying matter distribution in the universe. From earlier studies it is known that galaxies of different type cluster in different ways (Dressler 1980), which means that they do not trace the underlying mass exactly. Moreover, from the emptiness of large voids and the spikiness of the galaxy distribution with $\sim 100 \; h^{-1}$ Mpc spacing, specially at high redshifts, it is clear that if the structure has evolved according to standard gravitational instability theory then the galaxy distribution must be biased (Dekel & Lahav 1999). In addition, the distribution of galaxies obtained from semi-analytical models of galaxy formation does not trace the matter distribution from the corresponding N-body simulations, as is clearly shown in figure (1.1)

That means that it is mandatory to understand the process of galaxy biasing if one wants to constrain models of galaxy formation or to constrain the values of cosmological parameters from the observed distribution of galaxies.

Initially proposed by White & Rees (1978), the current framework of galaxy formation is divided into two parts: first, the dominant dark matter component in the universe collapses by gravitational instability into small lumps which then undergo a hierarchical process of formation of larger structures; second, the gas fraction trapped within the potential wells of the dark matter lumps cools down and condenses to form galaxies. The first stage of galaxy formation is driven by gravity alone, and therefore it seems easy to solve. However, the second stage is not yet well understood and many physical processes are involved in it. Nevertheless, these processes probably have their main influence on the individual properties of galaxies, and a negligible influence on the overall clustering properties.

Since the clustering properties of galaxies are mainly determined by the gravitational processes involved in the formation and clustering of dark matter haloes (i.e. virialized dark matter clumps), the clustering properties of the galaxy distribution can be fairly approached by studying the clustering properties of the dark matter haloes, where the only physical process involved is gravity.
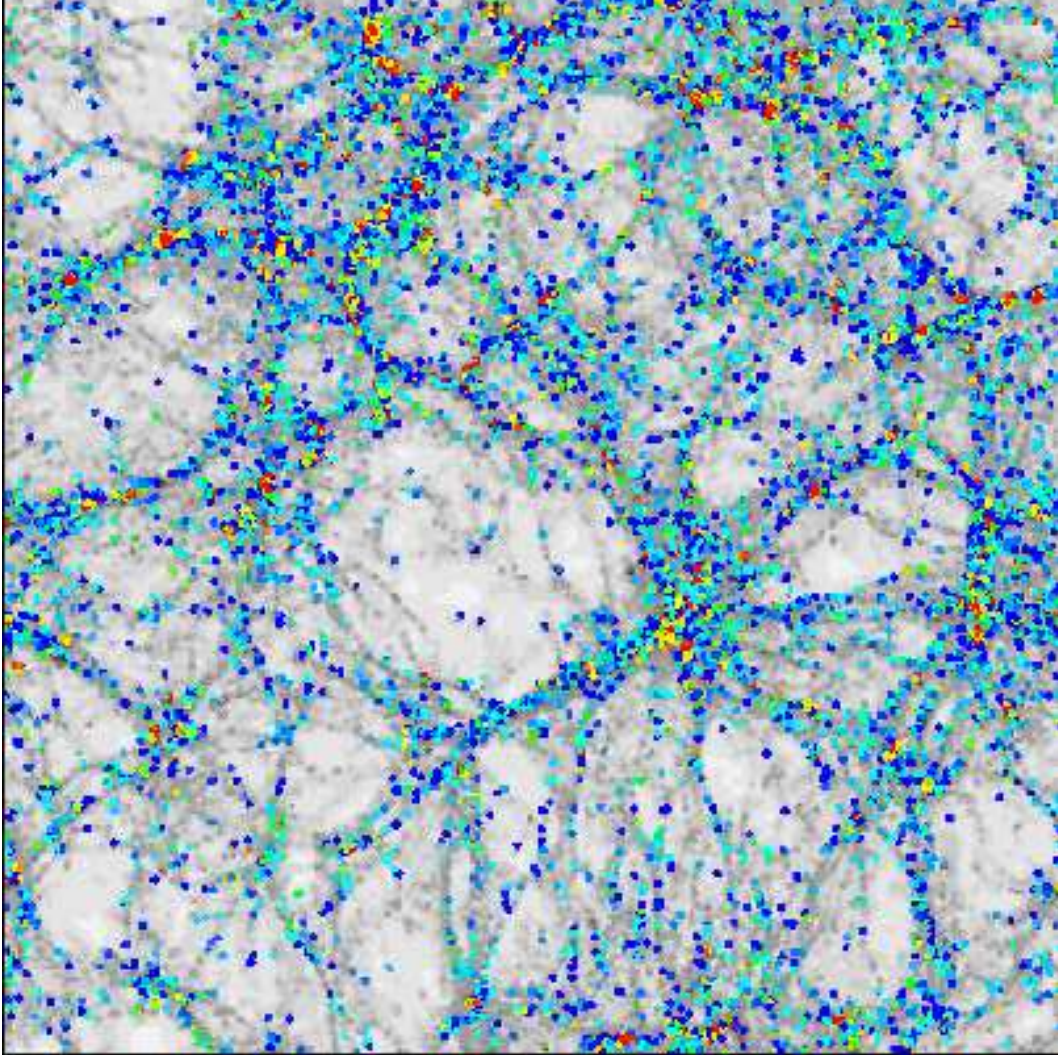
Figure 1.1: Present epoch population of galaxies (colored-circles), obtained from semi-analytical models of galaxy formation, superimposed to the dark-matter distribution (grey-colored) in the GIF τCDM N-body simulations. The picture corresponds to a slice of thickness 8 Mpc/h through the whole simulation box (85 Mpc/h on a side). The colors denote the B-V colors of the galaxies, ranging from red for ellipticals to blue for irregulars. Clearly, one can see that the distribution of galaxies does not trace the dark-matter distribution in a simple way. From Kauffmann et al. (1999)

.

In this thesis the relationship between the spatial distribution of dark matter haloes and the underlying dark matter distribution will be investigated. This relationship is commonly known as the *halo-mass bias*. In particular, some deterministic biasing models for the counts-in-cells moments, up to fourth order, of the spatial dark matter halo distribution, based on the spherical gravitational collapse and its ellipsoidal extension, will be tested. Moreover, the stochastic nature of the *halo-mass bias* relation will be analyzed in detail from the point of view of the conditional probability function, which specifies completely the bias relation in a statistical sense. In order to do that, a couple of theoretical distribution functions will be probed against the conditional probability obtained from high resolution N-body simulations and a model for the mean and the variance of the *halo-mass bias* relation will be probed against simulations. The implications of the stochasticity in *halo-mass bias* will be discussed.

The thesis is organized as follows. The remaining part of this chapter will briefly introduce the current framework of the cosmological model, composed by the definition of the background universe and the process of structure formation as the gravitational collapse of initial fluctuations from homogeneity, present in the primeval universe. A short introduction to some statistical measures of clustering is also given. Chapter 2 presents the test of some deterministic models of halo biasing against numerical N-body simulations. Chapter 3 presents the investigation of the stochastic nature of the *halo-mass bias* relation by searching for a good analytical descriptor for the conditional probability, among a couple of analytical distributions, and by testing a model for the mean and variance of the bias relation, based on spherical collapse. Additionally, in that chapter, the effect of the stochasticity in *halo-mass bias* is quantified. Finally, the main findings of this work will be summarized and discussed in chapter 4.

1.2   Cosmological Background

In the standard cosmological model it is assumed that the *Cosmological principle*, which states that the mass distribution in the universe is homogeneous and isotropic on large enough scales, is valid. An homogeneous and isotropic universe is the most simple model of the universe one can have. The space time of such a universe can be described by

means of the Robertson-Walker metric, which can be expressed as

$$ds^2 = c^2 \, dt^2 - a(t) \left(\frac{dr^2}{1 - kr} + r^2 d\Omega^2 \right), \tag{1.1}$$

where the spatial positions are denoted by the spherical coordinates (r, θ, ϕ). The squared solid angle element is $d\Omega^2 = d\theta^2 + \sin^2 \theta \, d\phi^2$. It should be noticed that, in these universes the spatial coordinates are comoving with the expansion of the universe, which is described by the dimensionless expansion parameter $a(t)$.

In general, the expansion parameter is defined to be unity at the present epoch $[a(t_0) = 1]$. The value of the parameter k defines the geometry of the universe under consideration to be flat $(k = 0)$, open $(k < 0)$ or closed $(k > 0)$.

Having the Robertson-Walker metric and assuming further that the matter and radiation content in the universe can be described as an ideal fluid, Einstein's field equations lead to the Friedmann equations

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right) + \frac{\Lambda c^2}{3} \tag{1.2}$$

and

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}. \tag{1.3}$$

Here G is the gravitational constant, ρ is the density of the matter-radiation fluid and p its pressure. The time derivative of the expansion factor is denoted by $\dot{a} \equiv \frac{d}{dt} a$.

Λ corresponds to the cosmological constant, also known as the energy density of vacuum.

For a complete description of the background universe it is necessary to provide the equation of state of the cosmic fluid. First of all, the cosmic fluid has three principal components: baryonic matter, dark matter, and radiation (relativistic massless particles). In the sense of the equation of state of the cosmic fluid, the history of the universe is commonly divided into two main epochs. The first epoch corresponds to the early universe, when radiation and relativistic particles were the dominant component of the energy density and the equation of state has the form $p = \frac{1}{3}\rho c^2$. As the universe expanded and cooled down, the energy density of the radiation decayed faster than the energy density of the non-relativistic matter. Thus, after a certain point in the history of the universe the energy density of the non-relativistic matter has started to dominate the cosmic energy density and so the pressure of the fluid can now be neglected. This epoch is known as the matter-dominated era.

The Friedmann equations together with the equation of state of the ideal cosmic fluid constitute a system of equations, whose solutions are characterized by some parameters. In the following I will introduce the definition of the parameters used to describe the Friedmann universes.

The first quantity to introduce is the Hubble constant, which is defined as the expansion rate of the universe at the present epoch (t_0)

$$H_0 = \left.\frac{\dot{a}}{a}\right|_{t_0}. \tag{1.4}$$

The Hubble constant is usually represented by the dimensionless factor h, defined by the expression $H_0 = 100\ h\ km\ s^{-1} Mpc^{-1}$.

Since the special case of a flat universe $(k = 0)$ is obtained for a special value of the matter density in the universe, it is natural to define a critical density ρ_c in terms of the Hubble constant

$$\rho_c = \frac{3H_0^2}{8\pi G}, \tag{1.5}$$

which is useful to define other dimensionless parameters, like the matter density parameter

$$\Omega_0 = \frac{\rho}{\rho_c}, \tag{1.6}$$

the vacuum density parameter

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2}, \tag{1.7}$$

and the curvature density parameter

$$\Omega_R = -\frac{kc^2}{a_0 H_0^2}. \tag{1.8}$$

All these quantities are defined for their values at the present epoch.

Therefore, a Friedmann universe can be described by the above defined cosmological parameters $(H_0, \Omega_0, \Omega_\Lambda, \Omega_r)$ and the expansion rate as a function of the expansion factor is given by

$$H^2(a) = H_0^2 \left[\Omega_\Lambda + \Omega_0 a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2}\right]. \tag{1.9}$$

This crucial relation is of great utility because it can be used to obtain the relation between time and the expansion factor

$$\frac{da}{dt} = a\ H_0 \left[\Omega_\Lambda + \Omega_0 a^{-3} + \Omega_r a^{-4} - (\Omega - 1)a^{-2}\right]^{1/2}. \tag{1.10}$$

In cosmology it is usual to express the time by means of the cosmological redshift. Due to the expansion of the universe, the electromagnetic radiation emitted with a wavelength λ_{em} is observed, at a distant place, having a wavelength $\lambda_{obs} > \lambda_{em}$. In other words, the electromagnetic signal has been redshifted by $z = (\lambda_{obs} - \lambda_{em})/\lambda_{em}$. This redshift is of particular importance and utility in cosmology because it can be directly measured from observations of spectra. If one neglects the peculiar motions the source can have, i.e. its deviations from the Hubble flow, the redshift can be related to the expansion factor by

$$z + 1 = \frac{1}{a_{em}}, \tag{1.11}$$

where a_{em} is the size of the universe at the emission time.

The exact values of the cosmological parameters are not known yet and their determination is still one of the major challenges of modern observational cosmology.

One of the simplest cosmological models corresponds to a flat universe with zero cosmological constant. This model is usually called Einstein-de-Sitter universe ($\Omega_0 = 1$, $\Omega_r = \Omega_\Lambda = 0$) and is a very attractive one from a theoretical point of view, since it comes out as a natural prediction from simple inflationary models.

The current state of the cosmic puzzle is still controversial. Observational evidence from the last years seems to be increasingly favoring a low mass universe with $\Omega_0 \sim 0.3$. Current observations are showing that maybe we are living in a vacuum dominated universe ($\Omega_\Lambda \sim 0.7$), with $\Omega_0 + \Omega_\Lambda = 1$.

1.3 Dynamics of Structure Formation

It is well known that the universe is populated, at large scales, by a wide range of structures, from small galaxies, to clusters of galaxies, super clusters and even to larger systems. Also it is known that there are very big regions of the universe without any galaxy (known as "voids"). This, altogether, is evidence for a universe far from homogeneity. From deep galaxy surveys we know too that the universe becomes statistically homogeneous at very large scales (e.g. at distances larger than $100 \; Mpc/h$). That means that, although the universe is inhomogeneous at small scales, if the cosmic fields are smoothed over large enough volumes, homogeneity is recovered and it is still possible to describe the global dynamics of the universe by the above introduced Friedmann models. Within this framework the structure in the universe is investigated as a deviation from homogeneity.

The most popular models of structure formation are based on a description of the background universe, a mechanism for generating small perturbations in the early universe, and a specification of the nature of the dark matter present in the universe.

The background universe is usually assumed to be a Friedmann model, completely specified by the value of the Hubble constant and the different components of the energy density. The existence of small perturbations in the energy density field at early epochs is explained by most of the models by means of processes which are assumed to have taken place in the very early universe (e.g. at $z \sim 10^{18}$ or earlier). So far, the most successful models in this regard are inflationary ones.

Let us assume a universe described by a Friedmann universe with small perturbations in the energy density field, specified by a power spectrum $P(k, z_{in})$ at very high redshift z_{in}. Since the perturbations at all relevant scales are small at epochs prior to the epoch when matter and radiation decoupled ($z_d = z \approx 1000$), it is possible to predict unambiguously the power spectrum $P(k, z_d)$ at the epoch of decoupling z_d using linear perturbation theory. In general the shape of the linearly evolved power spectrum of fluctuations at z_d depends on the nature of the dark matter present in the universe. In a universe dominated by "hot dark matter", i.e. particles which move relativistically when they decouple, large scales go non-linear first and smaller structures form by fragmentation. On the other hand, in a universe dominated by "cold dark matter", i.e. particles which move non-relativistically when they decouple and are moving very slowly at the present epoch, small scales go non-linear first and structure forms in a hierarchical way from the small to the large.

The evolution of the power spectrum after the decoupling epoch ($z < z_d$) is rather more complicated. In general, as long as the perturbations still small, the amplitude of the power spectrum evolves while the shape is preserved. For instance, in CDM universes the amplitude of the power spectrum grows equally for all scales between the epoch of decoupling and the present, assuming that linear theory is valid on all scales. This linearly evolved spectrum describes correctly the evolution of inhomogeneities at large scales but is not longer correct at small scales, due to the inherent non-linear nature of gravity, which is one of the major difficulties in understanding the physics at these small scales. Another difficulty is the need to understand the gas dynamics processes occurring at small scales.

Currently the most widely accepted cosmological models are those dominated by cold dark matter particles (CDM). Those models are specified by the relative energy density components. The structure content is determined by the power spectrum of fluctuations from homogeneity. The initial power spectrum of fluctuations in CDM universes can be described by (Efstathiou et al. 1992)

$$P(k) = \frac{Bk}{\{1 + [ak + (bk)^{3/2} + (ck)^2]^\nu\}^{2/\nu}}, \tag{1.12}$$

with

$$a = 6.4 \ \Gamma^{-1}h^{-1} \ kpc, \ b = 3.0 \ \Gamma^{-1}h^{-1} \ kpc, \ c = 1.7 \ \Gamma^{-1}h^{-1} \ kpc \ \text{and} \ \nu = 1.13.$$

Here Γ is the shape parameter of the power spectrum and its value is chosen to fit the power spectrum to a wide range of models. For the standard flat cold dark matter universe (SCDM) $\Gamma = h$; In the case of flat universes with a non-zero cosmological constant (ΛCDM) $\Gamma = \Omega_0 h$. For universes which are variants of the CDM model with more large-scale power due to decaying neutrinos (τCDM) the shape parameter is well approximated by $\Gamma \cong \Omega_0 h[0.861 + 3.8(m \ \tau)^{2/3}]^{1/2}$, where m is the neutrino mass (in units of 10 keV) and τ is its lifetime in years (Efstathiou et al. 1992). Within this framework the CDM cosmological model is completely specified by the values of the different components of the energy density and by the amplitude and shape of the power spectrum of fluctuations.

The growth of density fluctuations which are small ($|\delta| \ll 1$) at a given scale can be followed by linear perturbation theory. However, as the fluctuations grow, at some time $t_{nl}(\lambda)$ the density contrast at a given scale (λ) becomes comparable to unity. After this time the linear theory fails at this scale and, therefore, it is necessary to study the gravitational evolution in the non-linear regime. Since the complexity of the non-linear problem of gravitational collapse does not allow for general analytical solutions and only a few special cases with analytical solutions exist, like the spherical collapse model, numerical simulations of the gravitational dynamics of N bodies are necessary to get insights into the processes occurring within strong non-linear regions.

Analytically, the collapse of a spherical region of uniform density in an otherwise uniform background universe corresponds to one of the most simple models for the evolution of a gravitational perturbation one can have. This model is known as the "spherical collapse model" or as the "spherical top-hat collapse". It is known that the dynamics of

such a spherical perturbation are exactly the same as the dynamics of a closed universe with $\Omega_0 > 1$. The scale parameter of such universes obeys a cycloid evolution (i.e., the universe expands until a certain epoch and then recollapses due to its self gravitational potential.)

The evolution of the perturbation can be described as follows. Initially the perturbation expands along with the background universe. At a certain epoch the perturbation reaches its maximum radius (R_{max}) and its expansion separates out from the expansion of the background universe. After this epoch the perturbation undergoes a collapsing process.

If only gravity is acting on the perturbation, the sphere will collapse to a singularity. However this collapse does not occur in practice, since dissipative processes are usually present and therefore convert the kinetic energy of collapse into random motions. The final stage of the process will be thus a system which satisfies the virial theorem. That is, the internal kinetic energy of the self-gravitating system of masses will be equal to the half of its gravitational potential energy $2K = -U$. Within this framework it is straightforward to estimate the final size and density of the collapsed object.

Assuming that the total energy of the perturbation at its maximum expansion radius $R = R_{max}$ is in the form of gravitational potential energy, then from energy conservation, the kinetic energy when the perturbation has collapsed to half its maximum expansion radius is $K = -U/2$, which is the condition for equilibrium. At the time the perturbation has collapsed to half its maximum expansion $R = R_{max}/2$ the perturbation has increased its density by a factor of 8 while the background density has decreased, and the density of the perturbation at this time is about 170 times the background density. It is widely accepted that an object considered as virialized has a density larger than 200 times the background density.

1.3.1   The Press-Schechter Formalism

In Cold Dark Matter universes, only primordial fluctuations on very small scales survive after recombination and the nonlinear collapse of structures with sub-galactic size seems to be the first event to occur after recombination. These small scale-structures then cluster together in a "hierarchy", forming successively larger objects.

Press & Schechter (1974) have proposed an analytic formalism for the process of structure formation once the density perturbations have reached such an amplitude that they can be considered as having formed bound objects. The principal assumption in the Press-Schechter formalism is that, even if the field is nonlinear, the amplitude of large-wavelength modes in the final field will be close to that predicted from linear theory. It is known that a massive clump will undergo gravitational collapse if its average overdensity in a volume containing that mass exceeds some threshold of order unity (δ_c), independent of substructure. The properties of these bound structures can be estimated by an artificial smoothing of the initial density field using a filter function. If the filter function has some characteristic length R_f, then the typical size of the filtered fluctuations will be proportional to the characteristic length and, therefore, one can assign to them a mass $M \sim \rho_0 R_f^3$ [ρ_0 is the total background density of the model]. The exact form of the filter function is arbitrary and is generally adopted either as a Gaussian function or as a top-hat function (i.e. a sphere with uniform weight), due to analytical convenience.

For a Gaussian density field one has that the phases of the waves which make up the density distribution are random and the distribution of the amplitudes of the perturbations in a given smoothing volume V with characteristic scale $R \equiv R_f$ can be described by a Gaussian function

$$p(\delta) = \frac{1}{\sqrt{2\pi}\sigma(R)} \ \exp\left(-\frac{\delta^2}{2\sigma^2(R)}\right), \tag{1.14}$$

where $\sigma(R)$ is the linear *rms* in the smoothed version of δ. The probability that a given point lies in a spherical region of radius R with an overdensity larger than the critical overdensity for collapse ($\delta > \delta_c$) is

$$P(\delta > \delta_c \mid R) = \frac{1}{2}\left[1 - erf\left(\frac{\delta_c}{\sqrt{2}\ \sigma(R)}\right)\right], \tag{1.15}$$

Notice that the critical overdensity for collapse δ_c is given by a gravitational collapse model. For instance, in the spherical collapse model a density perturbation collapses when its linear overdensity reaches $\delta_c \sim 1.69$.

The Press-Schechter argument states that this probability is proportional to the probability that a given point has ever been part of a collapsed object with scale larger than R. That means that the only objects existing at a given epoch are those having only just reached the $\delta = \delta_c$ collapse threshold. For instance, if a point has an overdensity

larger than the critical overdensity for collapse, for a given scale R, then it will have an overdensity equal to the critical one ($\delta = \delta_c$) when smoothed at some larger scale and will be, therefore, counted as an object of the larger scale.

Thus, the fraction of the universe which has condensed into objects with mass $> M$ can be written in the universal form

$$F(> M) = 1 - erf\left(\frac{\nu}{\sqrt{2}}\right), \tag{1.16}$$

where $\nu = \delta_c/\sigma(M)$ corresponds to the threshold in units of the *rms* density fluctuations. There is a factor 2 with respect to equation (1.15) which was introduced by Press & Schechter (1974) to account for the problem that half of the mass remains unaccounted for, if using this probability.

One can express this integral probability in terms of the mass function $f(m)$ by

$$\frac{Mf(M)}{\rho_0} = \left|\frac{dF}{dM}\right|, \tag{1.17}$$

with the mass function defined in such way that $f(M)dM$ is the comoving number density of objects in the range dM; ρ_0 is the total comoving density.

Thus,

$$\begin{aligned}\frac{M^2\ f(M)}{\rho_0} &= \frac{dF}{d\ lnM}\\ &= \left|\frac{d\ ln\sigma}{d\ lnM}\right|\sqrt{\frac{2}{\pi}}\ \nu\ \exp\left(-\frac{\nu^2}{2}\right).\end{aligned} \tag{1.18}$$

$M^2f(M)/\rho_0$ is the multiplicity function, which corresponds to the fraction of the mass carried by objects in a unit range of lnM.

The Press-Schechter formalism and its extensions (Lacey & Cole 1994) are widely used to model the formation and evolution of structures in the universe, such as galaxy clusters and dark matter haloes, as well as to model the bias relation between the distribution of virialized objects and the underlying mass distribution.

1.4 Statistical Measures of Clustering

Now let us see how to test theories of structure formation using either observations or numerical simulations. In the following I shall briefly introduce some of the most widely used statistical measures of clustering. A detailed description of these statistics can be found in several excellent textbooks, like Peacock (1999) and Coles & Lucchin (1995).

The statistical measures most frequently used to study the large scale structure of the universe are: N-body correlation functions, which are defined to describe the clustering properties of a spatial distribution of objects; the power spectrum, which corresponds to the Fourier pair of the two-point correlation function; the void probability functions, which measures the probability of finding no objects in a randomly placed sphere of a given volume; and topological analysis, among others. Different statistical analysis measures different aspects of the observed clustering pattern. Statistics like the two-point correlation function, the cell-count variance and the galaxy power spectrum are directly related to the power spectrum of the fluctuations and therefore, can be used to constrain it. Methods like the higher-order correlations and fractal analysis can be used to study the role of self-similarity in the process of structure formation.

N-point Correlation Functions

One of the statistical quantities most widely used to measure the clustering in the universe corresponds to the correlation functions, which can be spatial three-dimensional measures or two-dimensional projections (angular correlation functions). The use of correlation functions was first suggested by Totsuji & Kihara (1969) and then continued in the 70's, mainly by Peebles. The correlation function is defined to describe the clustering properties of objects in space. These objects can be galaxies, clusters of galaxies or whatever entities, whose distribution in space is to be investigated.

The simplest correlation function corresponds to the two-point correlation function. It is widely used either in its spatial three-dimensional version or in its angular version. The spatial two-point correlation function is defined as the excess probability, in comparison with a random distribution, to find another object at a distance $\mathbf{r}_{12}$ from a given object,

$$dP = n_V^2[1 + \xi(\mathbf{r}_{12})]dV_1 \ dV_2, \tag{1.19}$$

where dP is the joint probability of finding one object in a small volume dV_1 and another one in the volume dV_2, separated by a distance $\mathbf{r}_{12}$. n_V is the mean number of objects per unit volume and the volume elements must be chosen randomly within a representative volume of the universe. $\xi(\mathbf{r}_{12})$ is called the spatial two-point correlation function. Due to statistical homogeneity and isotropy, ξ depends only on the modulus r of the separation vector $\mathbf{r}_{12}$.

The implementation of this definition, i.e. the way $\xi(r)$ can be measured in practice, requires the calculation of the expected number of pairs in a random distribution, within the sampling limits. This number is usually estimated creating a random catalog much larger in number than the sample under analysis. The correlation function is, then, estimated by counting pairs either within each catalog or between catalogues, giving several estimators of $1 + \xi$ as the ratio of different pair counts.

Count-in-Cells Statistics

A simple but very useful way to measure the clustering of objects (galaxies, haloes, clusters, etc.) on large scales corresponds to the distribution of count of objects in cells $P_V(N)$, defined as the probability of finding N objects in a randomly placed cell of volume V. In practice, rather than estimating the whole count-in-cells distribution (Probability Distribution Function), the first moments of the distribution function, like the variance, the skewness and the kurtosis, are estimated. Using only some of the moments of the count distribution leads to a loss of information in comparison with the use of the full distribution function. Nevertheless, the gain is a simple relationship between the moments of the count distribution and the correlation functions, because the cumulants $\bar{\xi}_Q(r)$ (or volume averaged correlation functions) of order Q can be expressed as a function of the central moments $\mu_Q \equiv \langle (\Delta N / \overline{N})^Q \rangle$ of the count-in-cells distribution, up to order Q; $\overline{N}$ is the mean number of objects in the cell. The cumulants $\bar{\xi}_Q(r)$ up to order fourth are written

$$\bar{\xi}_2(r) = \mu_2 - \frac{1}{\overline{N}}, \tag{1.20}$$

$$\bar{\xi}_3(r) = \mu_3 - 3\frac{\mu_2}{\overline{N}} + \frac{2}{\overline{N}^2}, \tag{1.21}$$

$$\bar{\xi}_4(r) = \mu_4 - 6\frac{\mu_3}{\overline{N}} - 3\mu_2^2 + 11\frac{\mu_2}{\overline{N}^2} - \frac{6}{\overline{N}^3}. \tag{1.22}$$

Here the additional terms, proportional to $\frac{1}{N}$, $\frac{1}{N^2}$, etc., at the right hand side correspond to a Poisson shot noise correction, due to discreteness effects. The cumulants are defined as a function of the correlation function by

$$\bar{\xi}_Q(R) \equiv \int dr_1...dr_j \ W_R(r_1)...W_R(r_j) \ \xi_j(r_1,...,r_j) , \qquad (1.23)$$

where $W_R(r)$ defines a suitable filter over a volume of size R.

The usual formulation for the moments of the count distribution is to define the quantities $S_Q \equiv \bar{\xi}_Q/\bar{\xi}_2^{Q-1}$. In the hierarchical model the S_Q's of the mass distribution should be constant, regardless of the size of the cell (Peebles 1980).

1.5 The Concept of Bias

The matter content in the universe is dominated by a dark component. It is currently known that the amount of luminous matter, i.e. matter emitting electromagnetic radiation, corresponds only to a small fraction of the total matter in the universe. Since galaxies are the building blocks of the "luminous" universe, they are the most suitable objects to be observed and studied to try to understand our universe. Therefore it is necessary first to understand the way galaxies trace the underlying dark matter field. The relationship between galaxies and the underlying total mass is called *galaxy-mass bias*. Thus, in order to be able to use the statistical properties of the observed galaxy distribution to understand better the physical processes of structure formation in the universe and to make use of galaxy catalogues to constrain the values of cosmological parameters, it is mandatory to have first a reliable picture of the *galaxy-mass bias* relation.

Currently, it is widely accepted that the cosmic structure we observe has been formed as the result of the growing of infinitesimal mass overdensities present in the primeval universe. Thus, as the universe evolves, these primeval seeds of the cosmic structure have undergone a growing process driven by gravity.

In 1982 Peebles introduced the term "Cold Dark Matter" to describe a population of exotic particles, whose existence was postulated in certain theoretical models of particle physics. The cold dark matter scenario mainly consists in assuming that the matter content in the universe is dominated by some form of collisionless dark matter.

White & Rees (1978) made a seminal proposal about the process of galaxy formation. They suggested that galaxy formation occurs in two different stages. At first they proposed that the dominant dark matter component of the universe collapses into small clumps at an early epoch and that these dark clumps continue clustering in a hierarchical way. This process leads to a self-similar distribution of bound masses, which correspond to what we usually call dark matter haloes. The second stage in the galaxy formation process corresponds then to the cooling and condensation of the baryonic mass trapped within the potential wells of the dark matter clumps.

The hydrodynamical part of the galaxy formation is still rather poorly understood and some effort has been spent in order to improve our knowledge about the physical processes involved and the role they play in the whole process of galaxy formation (e.g. White & Rees 1978; White & Frenk 1991).

Since the processes of formation and clustering of dark matter haloes in a cosmic density field involves only gravity, it appears that the problem of spatial galaxy clustering can be well approximated by understanding the spatial clustering of dark matter haloes and the formation of galaxies in individual dark haloes. In this way the clustering of dark matter haloes is studied by means of the gravitational theory and the formation of individual galaxies in a halo is traced using realistic models of galaxy formation in dark haloes that can now be constructed using semi-analytic models (e.g. Kauffmann et al. 1999; Cole et al. 2000; Somerville & Primack 1999). Indeed, there are quite a few recent investigations attempting to model galaxy clustering based on the halo scenario (e.g. Jing et al. 1998; Ma & Fry 2000; Scoccimarro et al. 2001; Peacock & Smith 2000; Seljak 2000; Sheth et al. 2001).

This also means that it is quite useful to approach the problem of galaxy clustering by studying the spatial clustering of dark matter haloes and assuming, initially, that the amount of galaxies hosted by a single halo is in general equal or very near to one. Of course this assumption might not be true, but it is a good initial starting point. In any case, the results from prior related studies and the results to be presented here show that this assumption is not far from reality.

Up to this point it is clear that it is worth to investigate the relationship between dark matter haloes and the mass density field (*halo-mass bias*). In this thesis, the statistical properties of the spatial dark halo distribution in cosmic density fields are studied.

Now let us introduce the concept of *halo-mass bias*. Let us define ρ and $\overline{\rho}$ as the mass density and the mean mass density, smoothed in regions of some given volume V, respectively. The mass density contrast (δ_m) in this volume is then defined as:

$$\delta_m \equiv \frac{\rho}{\overline{\rho}} - 1. \tag{1.24}$$

In the same way, if N_h and $\overline{N}_h$ correspond to the number of dark matter haloes and to the mean number of dark matter haloes in the volume V, respectively, the number density contrast of dark matter haloes (δ_h) is given by

$$\delta_h \equiv \frac{N_h}{\overline{N}_h} - 1. \tag{1.25}$$

The relationship between δ_h and δ_m is known as the halo-mass bias. A general way to represent this bias relation consists in expressing the halo number density contrast as a function of the mass density contrast

$$\delta_h(V) \equiv F(\delta_m(V)). \tag{1.26}$$

Since the mass and halo number density contrasts are defined in a given volume, the bias relation defined in equation 1.26 depends only on the masses within V and not on any mass-concentration outside of V. Therefore it is called "local halo-mass bias". The exact form of the function $F(\delta_m(V))$ is not known *a priori*. Several arguments and physical mechanisms for different kinds of biasing schemes have been proposed.

A simple approximation to the origin of biasing was formulated, for the case of galaxy biasing, by Kaiser (1984) and Bardeen et al. (1986) as a scheme for the biasing of high-density peaks in a Gaussian random field. In this scheme the two point spatial correlation functions of the mass (ξ_m) and of the galaxies (ξ_g) are related, in the linear regime, by means of the bias relation

$$\xi_g(r) = b^2 \xi_m(r), \tag{1.27}$$

where r is the scale at which the correlation is measured and b is a constant independent of the scale (i.e. the smoothing volume under consideration). This biasing scheme is called "linear bias". A more specific linear biasing scheme is usually adopted, where the local galaxy and mass density contrasts are related by

$$\delta_g(\mathbf{x}) = b\, \delta_m(\mathbf{x}), \tag{1.28}$$

where $\mathbf{x}$ is the spatial coordinate. From the definition of the two point correlation function, it can be shown that equation (1.27) follows from equation (1.28). This deterministic linear biasing scheme is probably too simplified.

More realistic models of biasing include deterministic models where b is a function of the scale and other variables. In the case of halo bias, for instance, a non-linear model of halo biasing has been developed by Mo & White (1996) based on the Press-Schechter formalism (Press & Schechter 1974) and its extensions (Lacey & Cole 1994). Their theoretical analysis provides a good approximation to the behavior of the non-linear halo biasing as a function of scale, time and halo mass. The use of N-body simulations has shown that the model proposed by Mo & White (1996) is a good approximation.

Nevertheless, since in general the number of haloes (and therefore of galaxies) formed in a given volume depends not only on the local mass density contrast but also on other properties of the mass distribution, the relation between the halo-number density field N and the local mass overdensity field δ_m is not expected to be deterministic. Thus, the *halo-mass bias* relation must be stochastic. In fact, the stochastic nature of the bias relation is already emphasized in the original paper of MW; in particular, MW pointed out that halo-exclusion can cause sub-Poisson variance, i.e. it is smaller than the mean. The effect of stochasticity may be important in the full distribution function of haloes (galaxies) as well as in the different biasing parameters needed to extract information from galaxy catalogues. Sheth & Lemson (1999) showed how the effects of stochasticity could be incorporated into the analysis of the higher-order moments of the halo distribution. Recently Somerville et al. (2001) used N-body simulations to study the stochasticity and non-linearity of the bias relation based on the formalism developed by Dekel & Lahav (1999). They analyzed the bias relation for haloes with masses larger than $1.0 \times 10^{12}~h^{-1}~M_\odot$ in spherical volumes of radius $8\,h^{-1}\mathrm{Mpc}$.

In this thesis the *halo-mass bias* relation will be investigated. In particular the validity of some deterministic biasing models, based on the spherical gravitational collapse and its ellipsoidal extension, will be tested against N-body simulations. Moreover, the stochastic nature of the *halo-mass bias* relation will be analyzed in detail from the point of view of the conditional probability function, which specifies completely the bias relation in a statistical sense. In order to do that, a couple of theoretical distribution functions will be probed against the conditional probability obtained from high resolution N-body

simulations and a model for the mean and the variance of the *halo-mass bias* relation will be probed against simulations. The implications of the stochasticity in *halo-mass bias* will be discussed.

Chapter 2

Deterministic Bias

2.1 Introduction

In the standard scenario of galaxy formation, it is assumed that galaxies form by the cooling and condensation of gas within dark matter haloes (e.g. White & Rees 1978; White & Frenk 1991). Thus, the problem of galaxy clustering in space can be approached by understanding the spatial distribution of dark matter haloes and galaxy formation in individual dark haloes. This approach is very useful for the following two reasons: (i) the formation and clustering properties of dark haloes can be modeled relatively reliably because of the simple physics involved (gravity only), (ii) realistic models of galaxy formation in dark haloes can now be constructed using either semi-analytic models (e.g. Kauffmann et al. 1999; Cole et al. 2000; Somerville & Primack 1999) or hydrodynamical simulations (e.g. Benson et al. 2001). Indeed, attempts have been made to use theoretical models of halo clustering to understand clustering properties of galaxies (e.g. Mo et al. 1997; Ma & Fry 2000; Scoccimarro et al. 2001; Peacock & Smith 2000; Seljak 2000; Sheth et al. 2001). Most of these investigations use the theoretical models presented in Mo & White (1996) and in Mo et al. (1997) (hereafter MJW) to calculate the second-order and high-order correlations of dark haloes. These models are based on the Press-Schechter formalism (Press & Schechter 1974) and its extensions (Lacey & Cole 1994).

The model prediction for the second moment, or the two-point correlation function, has been tested quite extensively by numerical simulations (Mo & White 1996; Mo et al. 1996; Jing 1998; Sheth & Tormen 1999; Governato et al. 1999; Colberg et al. 2000). The

results show that the model proposed by Mo & White works reasonably well over a large range of halo masses. However, a significant discrepancy between model and simulation results was found for low-mass haloes (Jing 1998; Sheth & Tormen 1999). Sheth et al. (2001) (hereafter SMT) suggested that the discrepancy at the low-mass end may be due to the fact that the model considered by Mo and White assumes spherical collapse for the halo formation while the collapse in realistic cosmological density fields may be better approximated by an ellipsoidal model. Indeed, SMT found that, if the ellipsoidal model is used, better agreement between the model and simulation results can be achieved in both the halo mass function and the two-point correlation function for low-mass haloes.

The performance of the MJW model for the high-order moments of the halo distribution has been tested in their original paper using scale-free N-body simulations with relatively low resolution. Although their results show that the theoretical model matches the simulation results, the limited dynamical range in the simulations used by them does not allow one to test the model for a large range of halo masses. Furthermore, although the MJW model has been extended to include ellipsoidal dynamics (Sheth et al. 2001), this extension has not yet been tested by simulation results.

In this chapter two sets of high-resolution Cold Dark Matter simulations are used to test the MW and MJW models and their ellipsoidal collapse extensions. One set has a very large simulation box (and so low mass resolution), which is used to control the finite-volume effect usually found in the analysis of high-order moments of the galaxy distribution (Colombi et al. 1994). The other set has a smaller simulation box but much higher mass resolution, which allows one to test the model for low-mass haloes.

The chapter is organized as follows: The procedure to obtain the moments of counts-in-Cells is presented in section (2.2). Section (2.3) introduces the theoretical models for the moments of halo counts-in-Cells. Analysis of the simulation data and the comparison of the theory with the simulation results are presented in section (2.4). An application of the theoretical model to the higher-order moments of Lyman-break galaxies is done in section (2.5). Finally, section (2.6) summarizes the results.

2.2 Moments of Counts-in-Cells

One of the most useful statistical quantities to measure clustering corresponds to the two-point correlation function, also known as the autocorrelation function. Nevertheless, the two-point correlation function does not contain complete information about the whole distribution of objects under consideration, and if one wants to look further, then it is necessary to make use of other statistical quantities, such as the N-point correlation functions (N=3,4,...).

In practice, however, the computation of the three- or four-point spatial correlation functions is quite expensive, in terms of the computation time and the computational resources needed. One way to overcome this problem is to use statistical quantities, which are easier to obtain and contain similar information as the N-point correlation functions. These quantities are the moments of the distribution of counts of objects, also known as the moments of counts-in-cells. They are related to the volume-average of the corresponding N-point correlation functions.

The distribution of the counts-in-cells is obtained, in practice, by counting the number of objects found within a randomly placed cell of a given volume and repeating the process sufficiently many times.

The calculation of the counts-in-cells moments of a discrete distribution of particles and the relation of such moments to the corresponding moments of the underlying continuous density field are described in detail in Peebles (1980). The relevant formulae are summarized in the following.

Let us assume that the cells used for the counts have spherical shape and therefore their volume is completely determined by their radii (R). The jth central moment of counts in cells of a point distribution is defined as

$$m_j(R) = \sum_{i=1}^{M} (N_i - \overline{N})^j, \tag{2.1}$$

where N_i is the number of particles counted in the ith sphere (cell), $\overline{N}$ is the mean number of counts: $\overline{N}(R) = \frac{1}{M} \sum_{i=1}^{M} N_i(R)$, and the summation is over the M sampling spheres. Notice that $\overline{N}$ is obtained directly from the counts.

The connected moments, μ_i, are defined through the central moments as

$$\mu_2 = m_2 \tag{2.2}$$

$$\mu_3 = m_3 \tag{2.3}$$

$$\mu_4 = m_4 - 3m_2^2 \,. \tag{2.4}$$

These relations are written up to the 4th order because these are the ones relevant for later discussion. For a point process, the shot noise also contributes to the quantities μ_j. These contributions become significant for small radii where the mean count $\overline{N}$ is small and should be properly subtracted. If the particle distribution is a Poisson sampling of the underlying density distribution, we can make the following subtractions to get the corrected connected moments:

$$k_2 = \mu_2 - \overline{N}, \tag{2.5}$$

$$k_3 = \mu_3 - 3\mu_2 + 2\overline{N}, \tag{2.6}$$

$$k_4 = \mu_4 - 6\mu_3 + 11\mu_2 - 6\overline{N}. \tag{2.7}$$

These quantities are related to the volume-averaged correlation functions by

$$\overline{N}^j \overline{\xi}_j = k_j \,, \tag{2.8}$$

where

$$\overline{\xi}_j = V_W^{-j} \int dr_1...dr_j \, W(r_1)...W(r_j) \, \xi_j(r_1, ..., r_j) \,, \tag{2.9}$$

and $W(r)$ is a top-hat spherical window with volume V_W.

2.3 Theoretical Models for the Moments of Dark Matter Halos

This section introduces the models for the moments of the spatial distribution of dark matter haloes, which are going to be tested against high-resolution numerical simulations.

The models to be tested correspond to the analytic model for the spatial clustering of dark matter haloes developed by Mo & White (1996) and the analytical model for the high-order correlations of haloes, both based on the spherical collapse model and the standard Press-Schechter formalism (Press & Schechter 1974) and its extension (Bond et al.

1991). These models are referred to as *spherical collapse based models.* In addition, the extension of the above mentioned models, based on ellipsoidal collapse model and its corresponding Press-Schechter extension, developed by Sheth et al. (2001) are also tested. These models shall be referred as *ellipsoidal collapse based models.*

2.3.1 Spherical Collapse Based Models

Two-Point Volume-Averaged Correlation Function

Let us start with the model for the variance of the spatial distribution of dark matter haloes in cosmic density fields, i.e., the second moment of the halo counts-in-cells. Mo & White (1996) (hereafter MW) have developed a model for the second-order correlations of dark matter haloes, based on the spherical collapse model and the Press-Schechter formalism (Press & Schechter 1974) and its extensions (Lacey & Cole 1994). The MW biasing model is summarized in the following. Without loss of generality, it is assumed an Einstein-de Sitter universe, i.e., $\Omega = 1$, $\Lambda = 0$.

Let us start with the spatial initial overdensity field $\delta(\vec{x}) \equiv [\rho(\vec{x}) - \overline{\rho}]/\overline{\rho}$, defined as the local deviation of the density field ($\rho(\vec{x})$) with respect to the mean density of the universe ($\overline{\rho}$). Let us assume that the spatial initial overdensity field $\delta(\vec{x})$ is Gaussian and therefore is described by a power spectrum $P(k)$. The field $\delta(\vec{x})$ is smoothed by convolving it with a spherically symmetric window function $W(r; R)$, with comoving characteristic radius R. Thus, the smoothed field is

$$\delta(\vec{x}; R) = \int W(|\vec{x} - \vec{y}| \; ; R) \; \delta(\vec{y}) \; d^3y, \tag{2.10}$$

which is equivalent to

$$\delta(\vec{x}; R) = \int \hat{W}(k; R) \; \delta_k \; e^{i\vec{k}\cdot\vec{x}} \; d^3k, \tag{2.11}$$

where δ_k and $\hat{W}(k; R)$ are the Fourier transforms of the spatial overdensity field $\delta(\vec{x})$ and the window function $W(r; R)$, respectively.

For a given window function the smoothed field of a Gaussian field is Gaussian too. Thus, since it is assumed that the spatial overdensity field $\delta(\vec{x})$ is Gaussian, the smoothed overdensity field $\delta(\vec{x}; R)$ is also Gaussian and, therefore, has the one point distribution

function:

$$p(\delta; R)d\delta = \frac{1}{(2\pi)^{1/2}} \ \exp\left(-\frac{\delta^2}{2\Delta^2(R)}\right) \ \frac{d\delta}{\Delta(R)}, \tag{2.12}$$

where $\Delta^2(R)$ is the rms mass fluctuation in a given window of comoving radius R:

$$\begin{aligned} \Delta^2(R) &= \ \left\langle [\delta(\vec{x}; R)]^2 \right\rangle \\ &= \ \int P(k)\hat{W}^2(k; R)d^3k. \end{aligned} \tag{2.13}$$

In linear perturbation theory δ and $\Delta(R)$ grow in the same manner, therefore, it is convenient to use their values linearly extrapolated to the present epoch. Keeping the notation as in MW, hereafter the formulae will be written in terms of extrapolated quantities, unless otherwise stated. The smoothing radius will not be explicitly written.

Throughout this work a top-hat window (filter) function is adopted, which has the form

$$W(r; R) = \frac{3V}{4\pi R^3} \quad (\text{for } r < R) \tag{2.14}$$

with Fourier transform

$$\hat{W}(k; R) = \frac{3(\sin kR - kR \ \cos kR)}{(kR)^3}. \tag{2.15}$$

The average mass contained in a Top-Hat window of radius R is simply

$$\overline{M}(R) = (4\pi/3)\overline{\rho}R^3. \tag{2.16}$$

For a certain power spectrum $P(k)$, a given spherical region (window) can be labeled equivalently by means of any one of the quantities R, Δ or M. They are equivalent variables.

Now, let us continue with the dark matter haloes field and its relation to the initial overdensity field. It is assumed that dark matter haloes are spherically symmetric, virialized dark matter clumps. From the spherical collapse model [see section 1.3] one has that a collapsing structure virializes (i.e., reaches its equilibrium state) at half its maximum radius of expansion, implying a density contrast (δ) at the time of collapse of about 178. Consider also, that in an Einstein-de Sitter universe a spherical perturbation with linear overdensity δ collapses at redshift $z_c = \delta/\delta_c$, where $\delta_c = 1.686$.

The mass M_1 of a halo is related to the initial comoving radius of the region from which it has been formed, by

$$M_1 = (4\pi/3)\overline{\rho}R_1^3. \qquad (2.17)$$

In what follows the properties of dark matter haloes will be labeled by the subscripts $1, 2, \ldots$, reserving the subscript 0 to label the properties of uncollapsed spherical regions.

In the Press-Schechter formalism the probability that a random mass element is part of a dark halo of mass exceeding M_1 at a given redshift z_1 is twice the probability that a surrounding sphere of mass M_1 in the initial conditions has a linearly extrapolated overdensity greater than δ_c at that redshift:

$$F(M_1, z_1) = \int_{(1+z_1)\delta_c}^{\infty} p(\delta; R_1)d\delta; \qquad (2.18)$$

$p(\delta; R_1)$ is given by equation (2.12). Another way to write this equation corresponds to the form:

$$F(M_1, z_1) = F(\nu_1) = erfc\left(\frac{\nu_1}{\sqrt{2}}\right), \qquad (2.19)$$

where the critical overdensity for collapse at redshift z_1 is defined as $\delta_1 = (1 + z_1)\delta_c$. $\nu_1 \equiv \delta_1/\Delta_1$ is the threshold in units of the rms density fluctuation and $erfc(x)$ is the complementary error function.

This probability is related to the mass distribution function $f(M_1, z_1)$. The comoving number density of objects in the mass range dM_1 is $f(M_1, z_1)dM_1$, and

$$\frac{M_1 f(M_1, z_1)}{\overline{\rho}} = \left|\frac{\partial F}{\partial M}\right|, \qquad (2.20)$$

where $\overline{\rho}$ is the current mean density of the universe. Therefore, the comoving number density of haloes, in current units, as a function of M_1 and z_1 is:

$$n(M_1, z_1)dM_1 = -\left(\frac{2}{\pi}\right)^{1/2}\frac{\overline{\rho}}{M_1}\frac{\delta_1}{\Delta_1}\frac{d\ln\Delta_1}{d\ln M_1}\exp\left(-\frac{\delta_1^2}{2\Delta_1^2}\right)\frac{dM_1}{M_1}. \qquad (2.21)$$

The relationship between halo abundances and the density field on larger scales has been modelled by MW as follows: Bond et al. (1991) derived the probability that the overdensity at a randomly chosen point is δ_0 when the initial field is smoothed on scale R_0 and does not exceed δ_1 for any larger smoothing scale. This is also considered, according

to the Bond et al. (1991) reinterpretation of the Press-Schechter formalism, to be the probability that a spherical region of initial radius R_0 has linear overdensity δ_0 and is not contained in a collapsed object of mass exceeding M_0 at redshift z_1 given by $\delta_1 = (1+z_1)\delta_c$. This probability is:

$$q(\delta_0, \delta_1; R_0) \, d\delta_0 =$$
$$\frac{1}{(2\pi)^{1/2}} \left[\exp\left(-\frac{\delta_0^2}{2\Delta_0^2}\right) - \exp\left(-\frac{(\delta_0 - 2\delta_1)^2}{2\Delta_0^2}\right) \right] \frac{d\delta_0}{\Delta_0}, \tag{2.22}$$

for $\delta_0 < \delta_1$ and zero otherwise. Notice that the subscript convention means that δ_0 refers to an uncollapsed region and therefore should be interpreted as the linear overdensity of that region extrapolated to the present, while δ_1, δ_2..., etc., refer to collapsed haloes and thus should be interpreted as $\delta_c(1 + z_i)$, with $z_i = z_1$, $z_2, \ldots$, etc being the redshift at which each halo is identified.

Bond et al. (1991) also show that the fraction of the mass in a region of initial radius R_0 and linear overdensity δ_0, which at redshift z_1 is contained in dark haloes of mass M_1 ($M_1 < M_0$ by definition) is given by

$$f(\Delta_1, \delta_1 \mid \Delta_0, \delta_0) \frac{d\Delta_1^2}{dM_1} dM_1 =$$
$$\frac{1}{(2\pi)^{1/2}} \frac{\delta_1 - \delta_0}{(\Delta_1^2 - \Delta_0^2)^{3/2}} \exp\left[-\frac{(\delta_1 - \delta_0)^2}{2(\Delta_1^2 - \Delta_0^2)} \right] \frac{\Delta_1^2}{dM_1} dM_1. \tag{2.23}$$

And thus the average number of M_1 haloes identified at redshift z_1 in a spherical region with comoving radius R_0 and overdensity δ_0 is

$$\mathcal{N}(1 \mid 0) \, dM_1 \equiv \frac{M_0}{M_1} \, f(1 \mid 0) \, \frac{d\Delta_1^2}{dM_1} \, dM_1, \tag{2.24}$$

where $f(1 \mid 0) \equiv f(\Delta_1, \delta_1 \mid \Delta_0, \delta_0)$. Since M_1 is identified as a collapsed halo at $z_1 > 0$ whereas M_0 is assumed to be uncollapsed at $z = 0$, we have $\delta_1 > \delta_0$.

From the analysis above it is clear that the number of haloes of mass M_1, identified at redshift z_1, which have formed from the matter initially contained within spheres of radius R_0 and linear overdensity δ_0 has a significant dependence on δ_0.

It is useful to quantify this dependence by calculating the average over-abundance of haloes in such spheres relative to the global mean halo abundance. This is:

$$\delta_h^L(1 \mid 0) = \frac{\mathcal{N}(1 \mid 0)}{n(M_1, z_1)V_0} - 1, \tag{2.25}$$

where $V_0 = (4\pi/3)R_0^3$ is the volume of the spherical region and the super-script L refers to Lagrangian quantities.

When the mass contained in the larger region is much greater than the mass of the haloes considered ($R_0 \gg R_1 \Leftrightarrow \Delta_0 \ll \Delta_1$ and $|\delta_0| \ll \delta_1$). Therefore

$$\delta_h^L(1 \mid 0) = b_0 \delta_0 = \frac{\nu_1^2 - 1}{\delta_1}\delta_0, \tag{2.26}$$

which means that the halo overdensity in these Lagrangian spheres is directly proportional to the linear mass overdensity. Notice that the constant of proportionality $b_0 = (\nu_1^2 - 1)/\delta_1$ is the same as the one obtained from the "peak-background split" argument by Efstathiou et al. (1988) and Cole & Kaiser (1989).

The next step is to model the clustering of dark matter haloes at recent epochs. For that MW have calculated the expected abundance of haloes in spheres that at the desired redshift z have radius R and (possibly) non-linear overdensity δ. They relate these quantities to the initial Lagrangian radius R_0 and the extrapolated linear overdensity δ_0 by means of a spherical collapse model. In this model each spherical shell moves as a unit and different shells do not cross until very shortly before they collapse through zero radius. Thus, the mass interior to each shell is constant, giving $R_0^3 = (1 + \delta)R^3$. Furthermore, since dark matter haloes are defined in the Press-Schechter formalism as objects identified at some given redshift, the mean number of haloes given by equation (2.24) can be taken as referring to haloes of mass M_1 identified at redshift z_1 within spheres of radius $R(R_0, \delta_0, z_1)$ and overdensity $\delta(\delta_0, z_1)$.

From the spherical collapse model one has that, for a spherical perturbation in an Einstein-de Sitter universe, the Eulerian radius R of a mass shell which had initial Lagrangian radius R_0 and mean linear overdensity δ_0 is given for $\delta_0 > 0$ by (see Peebles 1980, MW)

$$\frac{R(R_0, \delta_0, z)}{R_0} = \frac{3}{10}\frac{1 - \cos\theta}{|\delta_0|} \quad \text{and} \tag{2.27}$$

$$\frac{1}{1+z} = \frac{3 \times 6^{2/3}}{20} \frac{(\theta - \sin\theta)^{1/3}}{|\delta_0|}. \tag{2.28}$$

For $\delta_0 < 0$ one just has to replace $(1 - \cos\theta)$ by $(\cosh\theta - 1)$ in equation (2.27) and $(\theta - \sin\theta)$ by $(\sinh\theta - \theta)$ in equation (2.28).

Without loss of generality it can be assumed that $z = 0$ at the time when the clustering of haloes is examined. Therefore, δ_0 depends only on the present mass overdensity $\delta \equiv (R_0/R)^3 - 1$. For $|\delta| \ll 1$ $\delta_0(\delta)$ can be expanded in power series $\delta_0 = \sum_{k=0}^{\infty} a_k \delta^k$ (see Bernardeau 1992). Furthermore, MW introduced an interpolation formula that approximates accurately the relation between δ_0 and δ:

$$\begin{aligned}
\delta_0 &= -1.35(1+\delta)^{-2/3} + 0.78785(1+\delta)^{-0.58661} \\
&\quad -1.12431(1+\delta)^{-1/2} + 1.686647
\end{aligned} \tag{2.29}$$

Thus, MW have shown that, using the above mentioned assumption, the average overdensity of dark matter haloes in spheres with current radius R and current mass overdensity δ can be obtained from equations (2.21) and (2.24). It is

$$\delta_h(1 \mid 0) = \frac{\mathcal{N}(1 \mid 0)}{n(M_1, z_1)V} - 1, \tag{2.30}$$

where $V = (4\pi/3)R^3$, $R_0 = R(1+\delta)^{1/3}$ and δ_0 is determined from δ using the interpolation formula given in equation (2.29). When $R_0 \gg R_1$ and $|\delta_0| \ll \delta_1$ one has

$$\delta_h(1 \mid 0) \equiv b(M_1, z_1)\delta = \left(1 + \frac{\nu_1^2 - 1}{\delta_1}\right)\delta. \tag{2.31}$$

Notice that the halo overdensity is again directly proportional to the mass overdensity. In this case the constant of proportionality $b(M_1, z_1)$, which is commonly known as the *linear bias parameter*, is always positive. The first term in the definition of $b(M_1, z_1)$ comes from the contraction (or expansion) of the spherical region under analysis and the second term reflects the bias in the initial density field. MW have also shown that equation (2.31) is still valid for values of δ much greater than unity.

Section (2.4.2) shows the results of testing the MW model for the variance of haloes in counts-in-cells against numerical simulations using the relation

$$\sigma_h^2(M_1, z_1, R) = b^2(M_1, z_1)\sigma_m^2 \tag{2.32}$$

Higher-Order Moments of Counts-in-Cells

Mo et al. (1997) (MJW) have developed an analytical model for the hierarchical correlation amplitudes

$$S_{j,h}(R) = \overline{\xi}_{j,h}/\overline{\xi}_{2,h}^{j-1} \tag{2.33}$$

for $j = 3, 4, 5$ in the quasi-linear regime, where the subscript h stands for quantities of dark matter haloes. They have used the general formalism developed by Coles (1993) and Fry & Gaztanaga (1993).

In this model the statistical distribution of dark haloes within the initial density field, which is assumed to be Gaussian, is determined by an extension of the Press-Schechter formalism. The modifications of the distribution due to gravitationally induced motions are treated by means of a spherical collapse model (Mo & White 1996). The main results from this model, which are relevant for this analysis, are summarized here.

Following the same notation as in the last section, if the smoothed halo overdensity $(\delta_h(\vec{x}; R))$ is completely determined by the smoothed mass overdensity $(\delta(\vec{x}; R))$, then δ_h can be written as a function of δ, $\delta_g = f(\delta)$, independent of $\vec{x}$. Assuming $f(\delta)$ to be finite and smooth for δ around 0, the function f can be expanded in a Taylor series

$$\delta_h = f(\delta) = \sum_{k=0}^{\infty} \frac{b_k}{k!}\delta^k, \tag{2.34}$$

where the b_k are constants.

Fry & Gaztanaga (1993) have shown that if the j-point-volume-averaged mass correlation functions $(\overline{\xi}_j(R))$ have the hierarchical form

$$\overline{\xi}_j(R) = S_j\ \overline{\xi}_2^{j-1}(R), \tag{2.35}$$

then the transformation given by equation (2.34) preserves the hierarchical structure in the limit $\overline{\xi}_2(R) \ll 1$. Thus:

$$\overline{\xi}_{j,h}(R) = S_{j,h}\ \overline{\xi}_{2,h}^{j-1}(R). \tag{2.36}$$

Therefore, for the skewness and kurtosis ($j = 3$, 4) one has:

$$S_{3,h} = b^{-1}(S_3 + 3c_2), \tag{2.37}$$

$$S_{4,h} = b^{-2}(S_4 + 12c_2S_3 + 4c_3 + 12c_2^2), \tag{2.38}$$

where $c_k = b_k/b$, $b = b_1$ and the constants b_k are the coefficients in the expansion of the bias relation given in equation (2.34).

To obtain the coefficients b_k one follows the reasoning introduced in the last section, i.e., the MW model. Starting with the average overdensity of dark matter haloes of mass M_1 identified at redshift z_1 within spheres of radius $R(R_0, \delta_0, z_1)$ and overdensity $\delta(\delta_0, z_1)$, introduced in section (2.3.1):

$$\delta_h(1 \mid 0) = \frac{\mathcal{N}(1 \mid 0)}{n(M_1, z_1)V} - 1, \tag{2.39}$$

where $\mathcal{N}(1 \mid 0)$ is given by equation (2.24). Assuming that $R_0 \gg R_1$ one can replace $(\Delta_1^2 - \Delta_0^2)$ by (Δ_1^2) in the expression for $\mathcal{N}(1 \mid 0)$. Assuming further that $\delta \ll 1$, δ_h can be expanded in a Taylor series

$$\delta_h = \sum_{k=0}^{\infty} a_k \, \delta^k, \tag{2.40}$$

where the first coefficients are:

$$a_0 = 0, \quad a_1 = 1, \quad a_2 = -\frac{17}{21}, \quad a_3 = \frac{341}{567}$$

MJW obtained the coefficients b_k for a halo with mass M_1 corresponding to a linear overdensity δ_1, which collapses at redshift $z_1 = \delta_1/\delta_c - 1$ (with the critical overdensity for spherical collapse being $\delta_c = 1.686$),

$$b_1 = 1 + \frac{\nu_1^2 - 1}{\delta_1}, \tag{2.42}$$

$$b_2 = 2(1 + a_2)\frac{\nu_1^2 - 1}{\delta_1} + \left(\frac{\nu_1}{\delta_1}\right)^2 (\nu_1^2 - 3), \tag{2.43}$$

$$b_3 = 6(a_2 + a_3)\frac{\nu_1^2 - 1}{\delta_1} + 3(1 + 2a_2)\left(\frac{\nu_1}{\delta_1}\right)^2 (\nu_1^2 - 3)$$

$$+ \left(\frac{\nu_1}{\delta_1}\right)^2 \frac{\nu_1^4 - 6\nu_1^2 + 3}{\delta_1}, \tag{2.44}$$

where $\nu_1 \equiv \delta_1/\sigma(M_1)$ [with $\sigma(M_1)$ being the rms of the density fluctuation given by the density spectrum linearly extrapolated to the present time].

The bias coefficients are given for the present-day descendants (at redshift $z_0 = 0$) of haloes identified at redshift z_1. The formalism can be easily extended to the case where $z_1 > z_0 > 0$. In this case, we replace δ_1 by $\delta_1 \, D(z_0)/D(0)$ (where $D(z)$ is the linear growth rate evaluated at redshift z) while keeping ν_1 unchanged.

2.3.2 Ellipsoidal Collapse Based Model

The mass function predicted by the Press-Schechter model in its standard form (i.e., assuming spherical collapse) is reasonably accurate at the high mass end. Nevertheless, it has more low mass objects than are found in numerical simulations of hierarchical clustering (Lacey & Cole 1994; Sheth & Tormen 1999).Sheth et al. (2001) (hereafter SMT) argue that this can be because the spherical collapse approximation to the dynamics may not be accurate, since Gaussian density fields are inherently triaxial (Bardeen et al. 1986). SMT modified the standard Press-Schechter formalism by incorporating the effects of non-spherical collapse.

In their model SMT assumed that bound structures form from an ellipsoidal rather than a spherical collapse. In the spherical collapse the value of the critical overdensity for collapse δ_c is independent of the initial size of the region from which the halo is formed and, thus is also independent of the final mass of the object. The main effect of including the dynamics of ellipsoidal rather than spherical collapse is to introduce a simple dependence of the critical overdensity required for collapse on the halo mass ($\delta_{ec} = \delta_{ec}(M)$). For the ellipsoidal collapse model it is assumed that the collapse of a region depends on the surrounding shear field and on its initial overdensity and provide a fitting function for the relation between the overdensity value required for collapse and the mass of the final object M. For that they have used the model for the gravitational collapse of homogeneous ellipsoids described by Bond & Myers (1996), where the evolution of the density perturbation is assumed to be better described by the initial shear field than by the initial density field, the Zeldovich approximation in the linear regime is recovered choosing initial conditions and external tides, and the time of virialization of the object is defined as the time when the third axis collapses, according to the prescription that collapse along each axis is frozen once it has shrunk by some critical factor; this freeze-out radius is chosen so that the density contrast at virialization time is the same as in the spherical collapse models (i.e., ≈ 178). SMT state that their results are not very sensitive to the exact value of the freeze-out radius.

Considering the collapse of ellipsoids from an initially Gaussian fluctuations field SMT have estimated the relation between δ_{ec} and the mass of the halo:

$$\delta_{ec}(\Delta, z) = \delta_c(z) \left(1 + \beta \left[\frac{\Delta^2(z)}{\Delta_\star^2(z)} \right]^\gamma \right), \tag{2.45}$$

where $\Delta_\star(z) \equiv \delta_c(z)$, $\beta = 0.47$ and $\gamma = 0.615$. Notice that the power spectrum enters only in the relation between the mass M and $\Delta(M)$, and cosmology enters only in the relation between the critical overdensity for collapse in the spherical model and the redshift z.

SMT have shown also that equation (2.45) implies that for massive objects (i.e., $\Delta/\Delta_\star \ll 1$) $\delta_{ec}(\Delta, z) \cong \delta_c(z)$ and that $\delta_{ec}(\Delta, z)$ increases with $\Delta(M)$, so it is larger for less massive objects. Therefore massive objects are well described by the spherical collapse model, whereas smaller objects are more influenced by external tides and must have a greater internal density to be able to hold themselves as they collapse.

SMT have used equation (2.45) to include the effects of ellipsoidal collapse into the (Bond et al. 1991) excursion set model and therefore to obtain an estimate of the mass function associated with ellipsoidal collapse:

$$\nu f(\nu) = 2A \left(1 + \frac{1}{\nu^{2q}} \right) \left(\frac{\nu^2}{2\pi} \right)^2 \exp \left(-\frac{\nu^2}{2} \right), \tag{2.46}$$

where $q = 0.3$ and $A = 0.322$. In the case of spherical collapse one has that $q = 0$ and $A = 1/2$.

If the ellipsoidal model is used to define collapsed haloes, the coefficients b_k of the bias relations for the first moments of counts-in-cells take the following form:

$$b_1 = 1 + \epsilon_1 + E_1, \tag{2.47}$$

$$b_2 = 2(1 + a_2)(\epsilon_1 + E_1) + \epsilon_2 + E_2, \tag{2.48}$$

$$b_3 = 6(a_2 + a_3)(\epsilon_1 + E_1) + 3(1 + 2a_2)(\epsilon_2 + E_2) + \epsilon_3 + E_3, \tag{2.49}$$

where

$$\epsilon_1 = \frac{\alpha\nu^2 - 1}{\delta_1}, \tag{2.50}$$

$$\epsilon_2 = \frac{\alpha\nu^2}{\delta_1^2}(\alpha\nu^2 - 3), \tag{2.51}$$

$$\epsilon_3 = \frac{\alpha\nu^2}{\delta_1^3}(\alpha^2\nu^4 - 6\alpha\nu^2 + 3), \tag{2.52}$$

$$E_1 = \frac{2p/\delta_1}{1 + (\alpha\nu^2)^p}, \tag{2.53}$$

$$\frac{E_2}{E_1} = (\frac{1 + 2p}{\delta_1} + 2\epsilon_1), \tag{2.54}$$

$$\frac{E_3}{E_1} = \frac{4(p^2 - 1) + 6p\alpha\nu^2}{\delta_1^2} + 3\epsilon_1^2, \tag{2.55}$$

and $\alpha = 0.707$, $p = 0.3$. These formulae reduce to the original MJW model for $\alpha = 1$ and $p = 0$.

Using the $b = b_{k's}$ from the spherical collapse based models [equations (2.42)–(2.44)] and from the ellipsoidal collapse based model [equations (2.47)–(2.49)] in equations (2.32), (2.37) and (2.38) and taking S_3 and S_4 in these equations to be the skewness and kurtosis of the mass distribution measured directly from the N-Body simulations, we can calculate the variance, skewness and kurtosis for the distribution of dark haloes as predicted by the MW and the MJW models and its SMT extension, respectively.

2.4 Test by N-body Simulations

2.4.1 Simulations

In the present analysis we use two sets of cosmological N-body simulations, which have been obtained as part of the VIRGO (Jenkins et al. 1998) and the GIF (Kauffmann et al. 1999) projects. These two sets of simulations differ in the size of the simulation boxes and in the mass resolution, with the VIRGO simulations having a larger simulation box and lower mass resolution than the GIF ones. From the VIRGO simulations we have analyzed the ΛCDM model in order to test the models in a volume large enough so that the effects due to the finite sampling volume may be negligible (see below). We compare the results with those obtained from the GIF simulations to see how comparisons between models and simulations can be made for simulations with a relatively small volume. For the GIF simulations, we focus on the τCDM and ΛCDM models. The parameters characterizing the simulations are summarized in table 1. Further details can be found in Kauffmann et al. (1999) and Jenkins et al. (1998).

For each simulation there are several output files corresponding to different evolutionary times (redshifts) and for each of these output times there is a halo catalog containing information about haloes identified using the friends-of-friends group-finder algorithm

Model	Ω_0	Ω_Λ	h	σ_8	Γ	Box Size [Mpc/h]	$N_{\rm p}$	$m_p/M_\odot h^{-1}$
GIF-τCDM	1.0	0.0	0.5	0.6	0.21	85	256^3	1.0×10^{10}
GIF-ΛCDM	0.3	0.7	0.7	0.9	0.21	141	256^3	1.4×10^{10}
VIRGO-ΛCDM	0.3	0.7	0.7	0.9	0.21	239.5	256^3	6.86×10^{10}

Table 2.1: Parameters characterizing the simulations used in the analysis. Ω_0 and Ω_Λ are the density parameters for matter and for the cosmological constant, respectively, h is the Hubble parameter, σ_8 is the rms of the density field fluctuations in spheres of radius $8\,h^{-1}$ Mpc, and Γ is the shape parameter of the power spectrum. Also given are the size of the simulation box, the total number of particles and the mass per dark matter particle in a simulation.

with a linking length 0.2 times the mean interparticle separation. Only haloes containing 10 or more particles are included in the halo catalogues. The physical quantities available from each of these halo catalogues are: the index of the most-bound particle in the halo, which corresponds to the position of the halo as well as the central 'galaxy' within it; the virial radius (R_{vir}), defined as the radius (from the central particle) within which the overdensity of dark matter is 200 times the critical density; the virial mass (M_{vir}), which is the mass (or, equivalently, the total number) of dark matter particles within the virial radius; the circular velocity [$V_c = (GM_{vir}/R_{vir})^{1/2}$].

We have also generated several catalogues of the present-day positions of the central objects corresponding to the most-bound particles in haloes identified at an earlier epoch. These catalogues might be interpreted as 'galaxy catalogues' if we assume that the positions of galaxies at the present epoch correspond to those of the central particles within virialized objects identified at high redshifts. This concept is related to the assumption in models of galaxy formation that galaxies form by the cooling and condensation of gas within dark matter haloes (White & Frenk 1991). However, this interpretation does not take into account subsequent galaxy mergers.

We apply the counts-in-cells analysis described in the last section to the mass distributions and halo catalogues. To do this, we place spheres in a regular mesh of 30^3 centers and count the number of objects at each center over a set of concentrical spheres, which allows us to compute the desired statistical quantities at different radii.

2.4.2 Testing the models for the variance

The volume-averaged two-point correlation functions have been obtained from the mass distribution and from the several halo samples in the simulations. For that, the procedure described in section (2.2) has been followed.

The analysis has been performed for two different cases. In the first, the high-order moments are calculated at the same time when the dark haloes are identified. In the second, haloes are identified at some high redshift while the calculations of the high-order moments are performed for their descendants at a later time. In all cases, the redshift at which halo identification is made is denoted by z_1, while the redshift at which the high-order moments are calculated is denoted by z_0.

Figures (2.1)–(2.3) show the variance from the VIRGO ΛCDM simulation, together with the predictions from the MW and its ellipsoidal collapse extension. Both the MW model and the SMT ellipsoidal collapse extension work remarkably good in all cases. In these two figures the prediction of the MW model with $\overline{\xi}_2$ given by the perturbation theory (see Bernardeau 1994) is also plotted. The fact that this prediction also matches the simulation results suggests that the moments obtained from the VIRGO simulations are not affected significantly by the finite-volume effect and confirms that the MW model is a good approximation to the second-order moment of haloes that are not much smaller than $M_\star$ [defined by $\sigma(M_\star) = 1.68$]. Similarly, Figures (2.4)–(2.6) show the variance from the GIF ΛCDM simulation, together with the predictions from the MW and its ellipsoidal collapse extension.

With their high mass resolutions, GIF simulations allow one to test the theoretical models for haloes with mass $M \ll M_\star$. Since the GIF simulations have relatively small simulation boxes, the moments are expected to be affected by the finite-volume effect (Colombi et al. 1994). Nevertheless the finite-volume effect on the variance is expected to be negligible. In any case, this effect in each simulation is expected to be similar for both the mass distribution and the halo distribution. Thus, to test the bias model given in equation (2.32) by a numerical simulation we should use the value of $\overline{\xi}_2$ obtained directly from the simulation, because it is the simulated power spectrum (not the theoretical spectrum) that is responsible for the clustering in the simulation. Figure (2.7) shows the results obtained for the GIF simulations for haloes identified and analyzed at the present epoch. As one can see, there is good agreement between model predictions and simulation

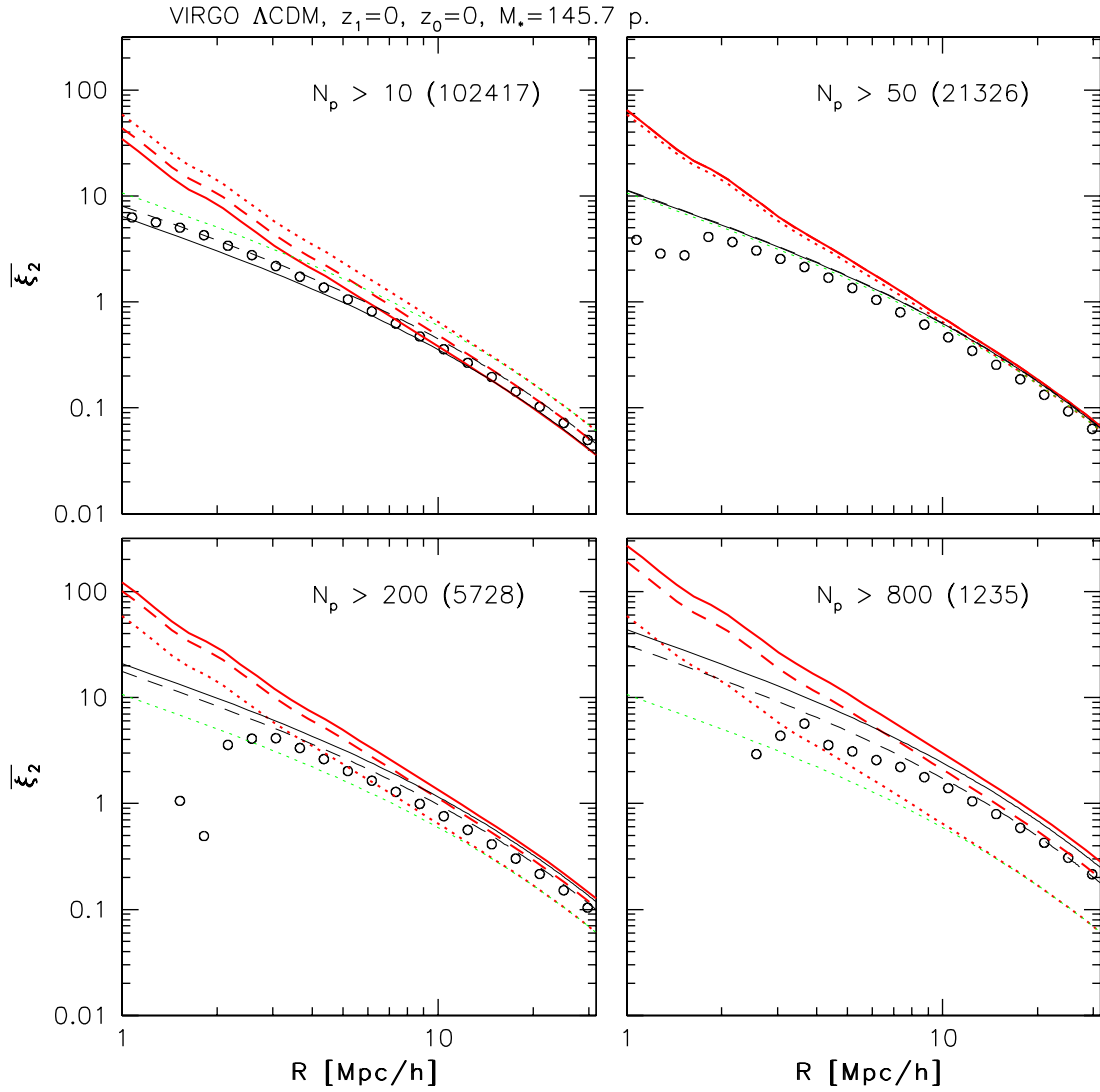results. For haloes with masses much smaller than $M_\star$, the SMT model gives a better fit than the MW model.
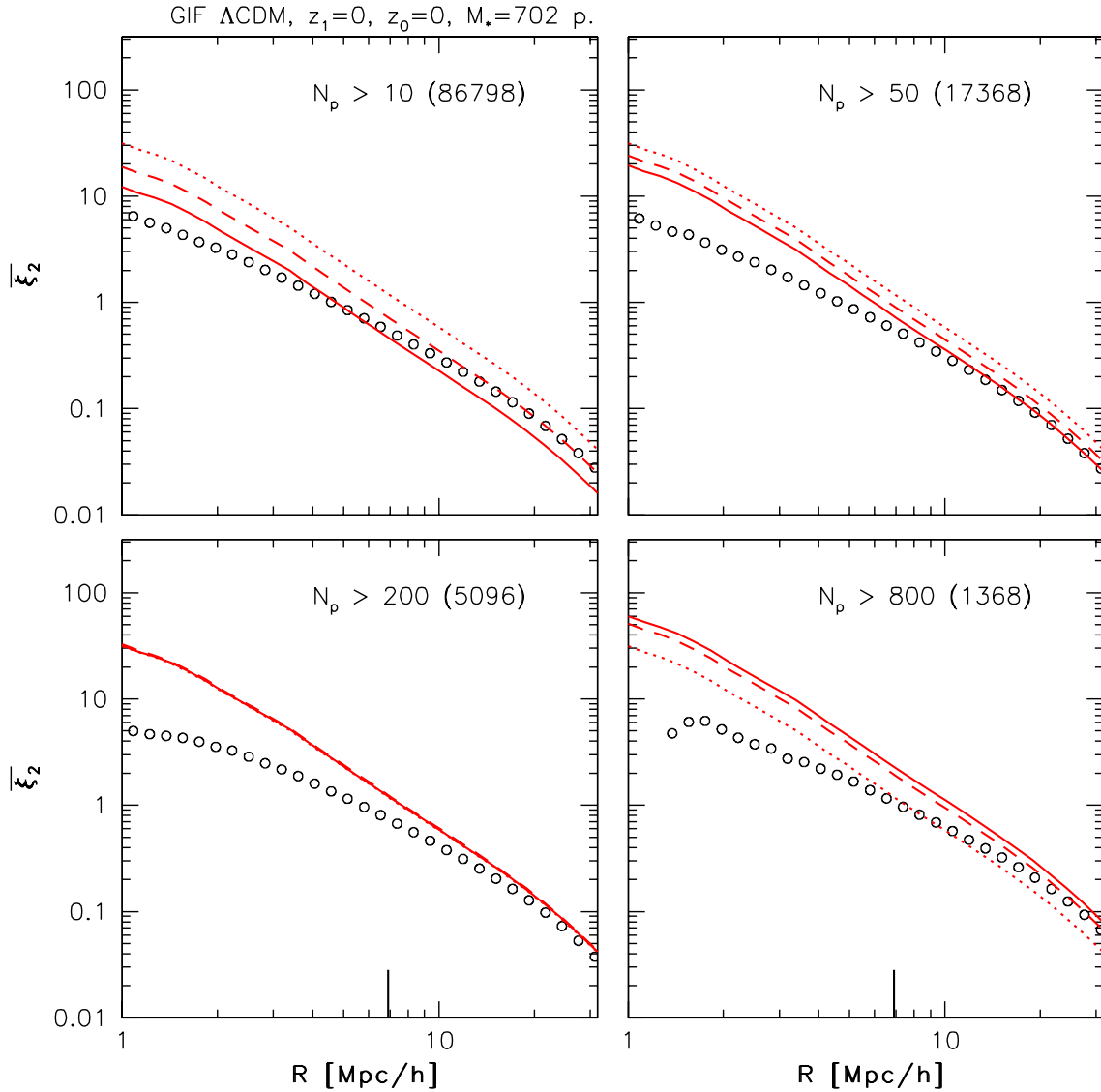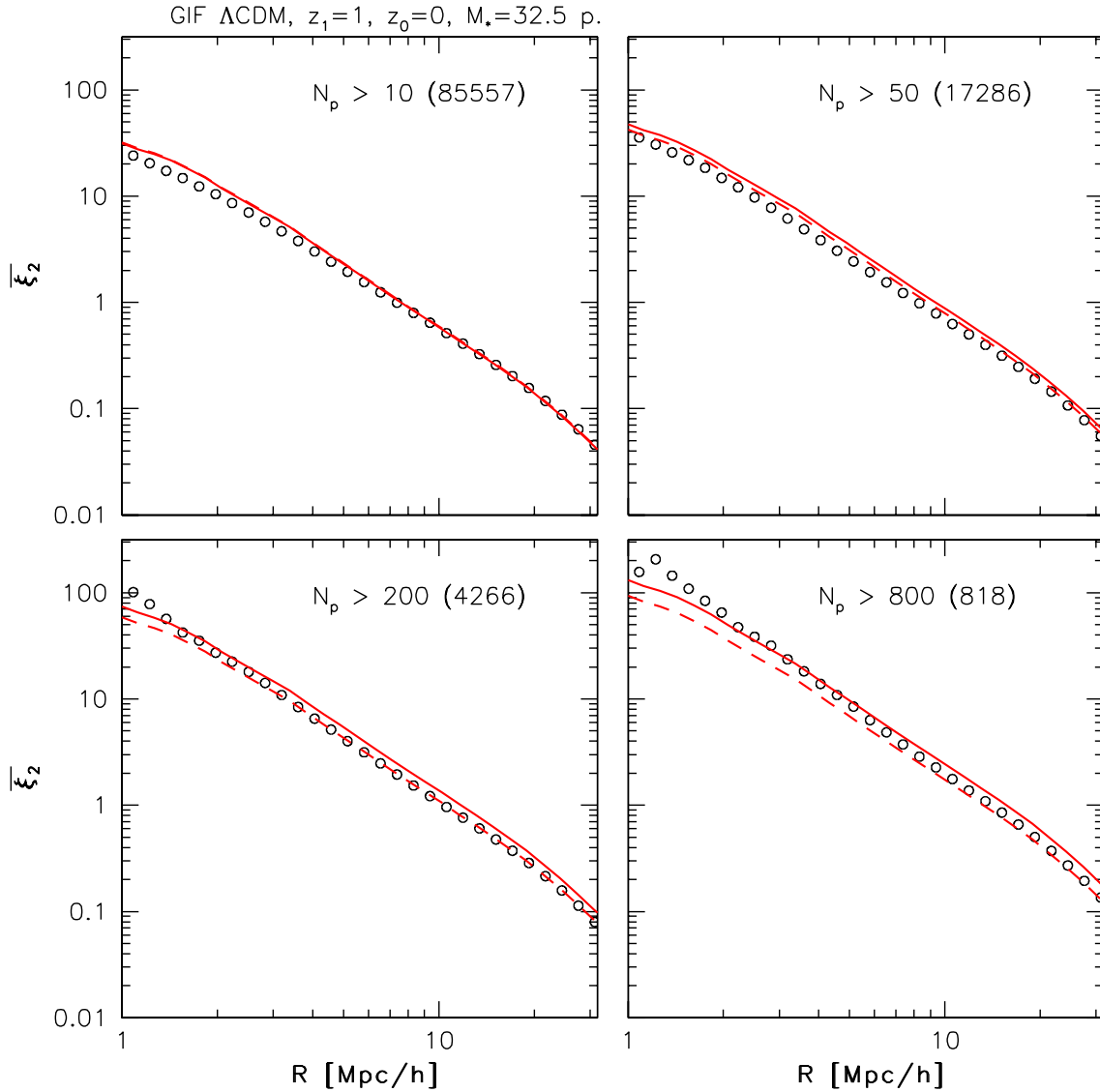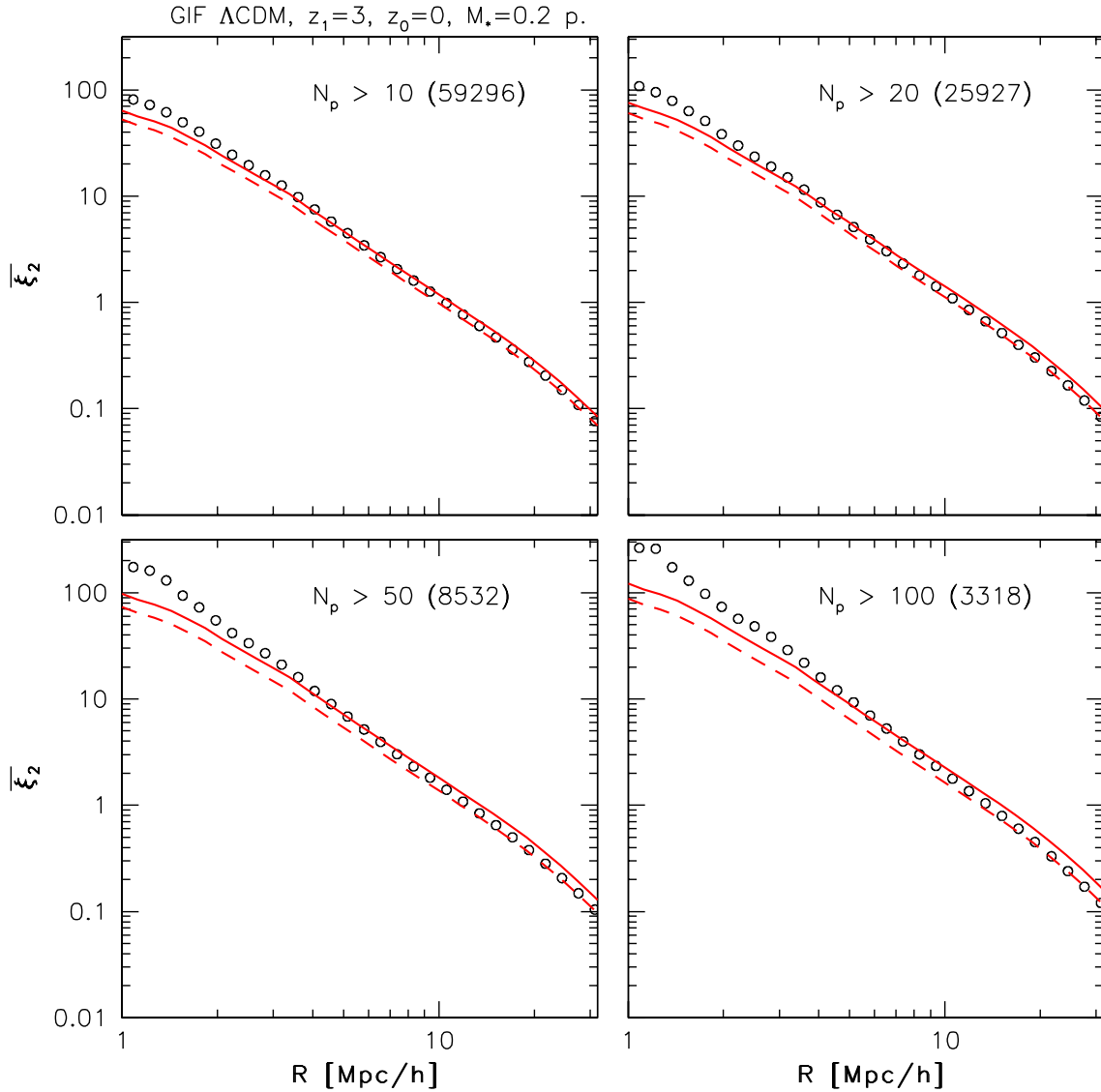

Figure 2.1: Variance $\overline{\xi}_2$ of dark haloes with different mass ranges obtained from the counts-in-cells analysis (symbols), from applying the bias model from MW (solid line) and its SMT extension (dashed-line). The moments for the mass distribution are shown by dotted lines. Thin lines correspond to quantities obtained using the variance of the mass given by perturbation theory (Bernardeau 1994), whereas thick lines correspond to quantities obtained using the variance of the mass directly from the simulations. Results are shown for the VIRGO ΛCDM simulations. The haloes have been identified and analyzed as indicated in the plot. The value of M_* is also written for more information. Each box corresponds to a different range of masses of haloes. The quantities in parenthesis correspond to the number of haloes in each sample.
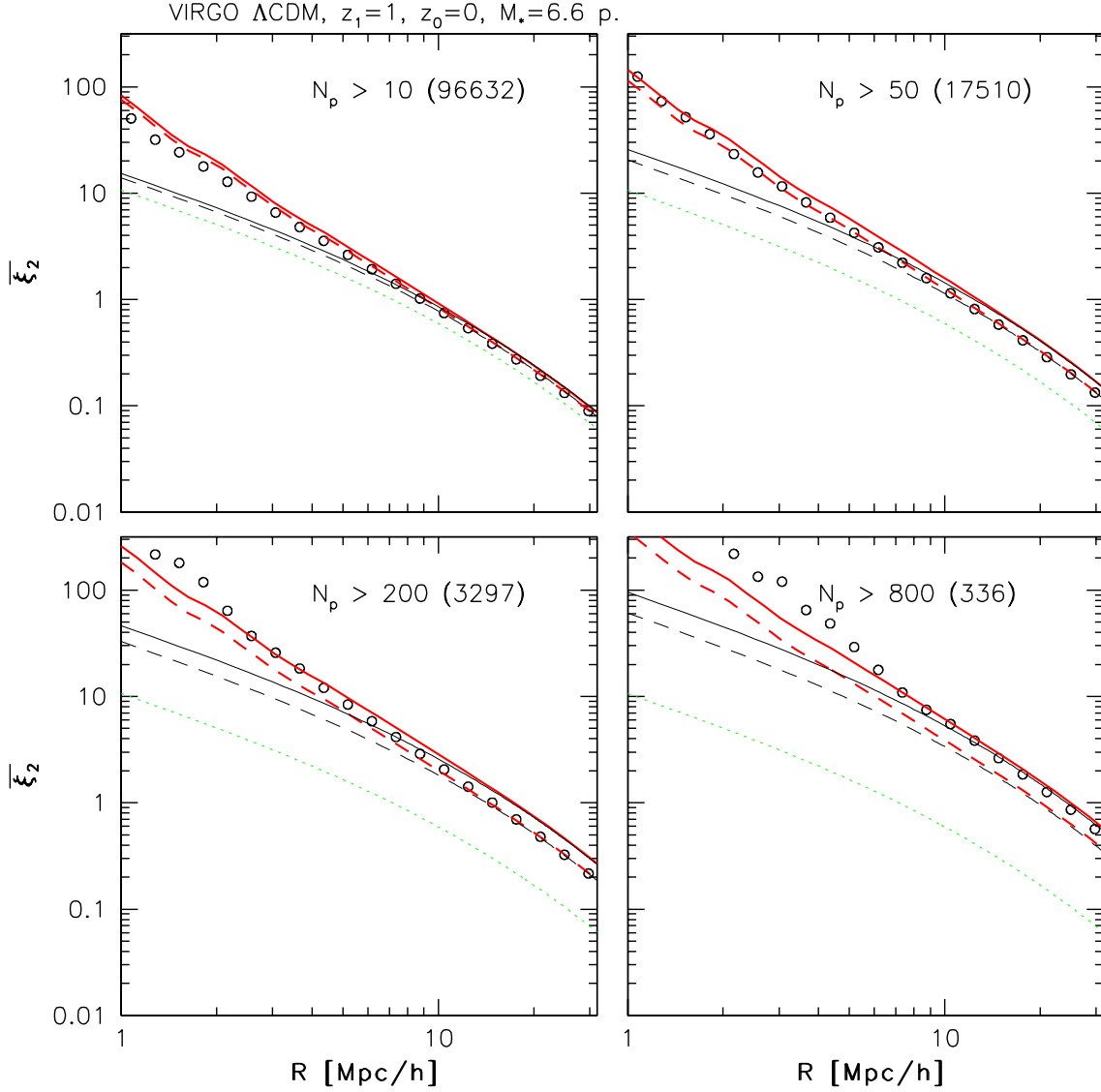
Figure 2.2: Variance $\overline{\xi}_2$ of haloes in the VIRGO ΛCDM simulations for haloes identified at $z = 1$ and analyzed at the present time. The lines and symbols, as well as the notation have the same meaning as in figure 2.1.
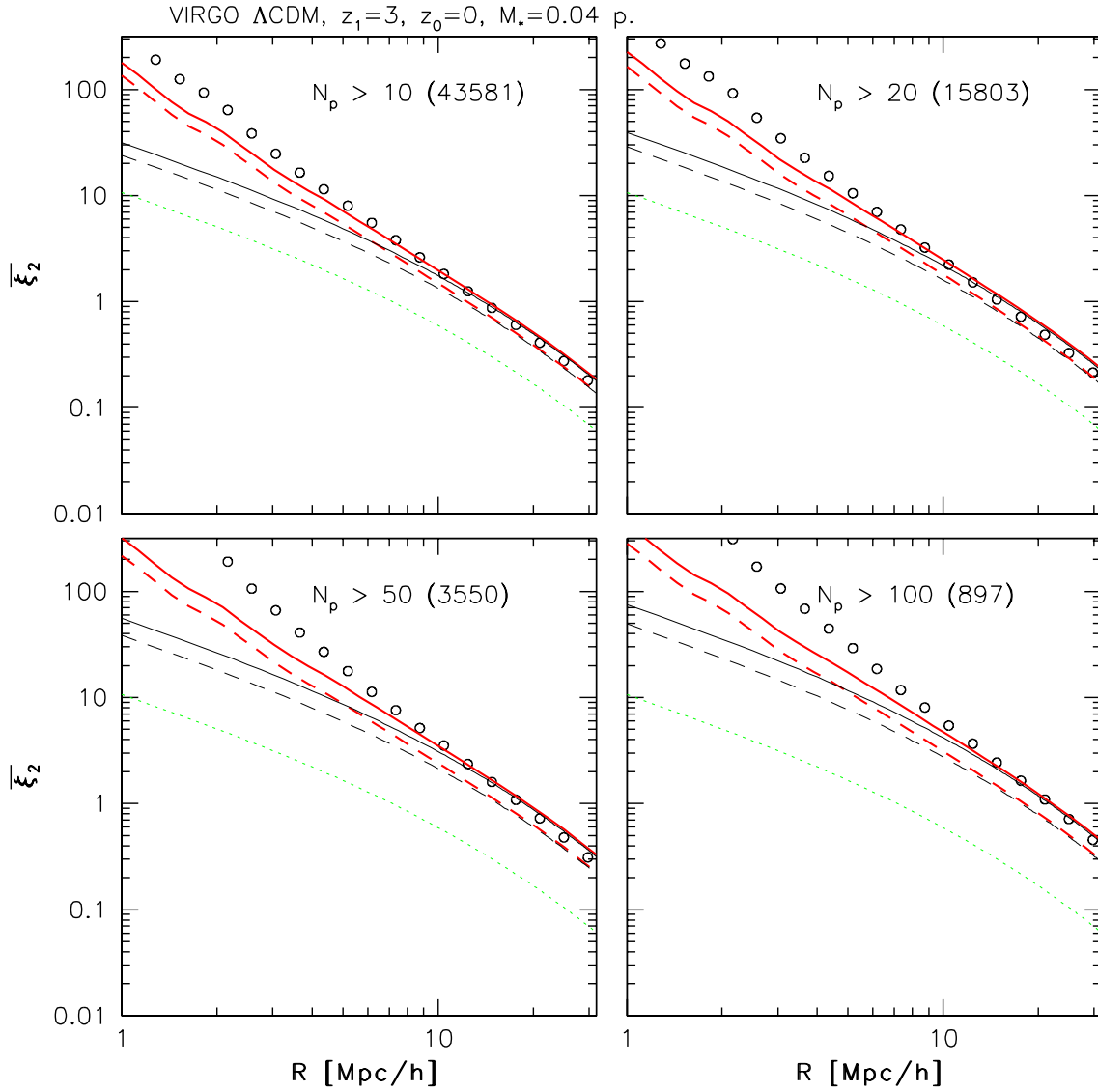
Figure 2.3: Variance $\overline{\xi}_2$ of haloes in the VIRGO ΛCDM simulations for haloes identified at $z = 3$ and analyzed at the present time. Lines, symbols and notation have the same meaning as in figure 2.1.
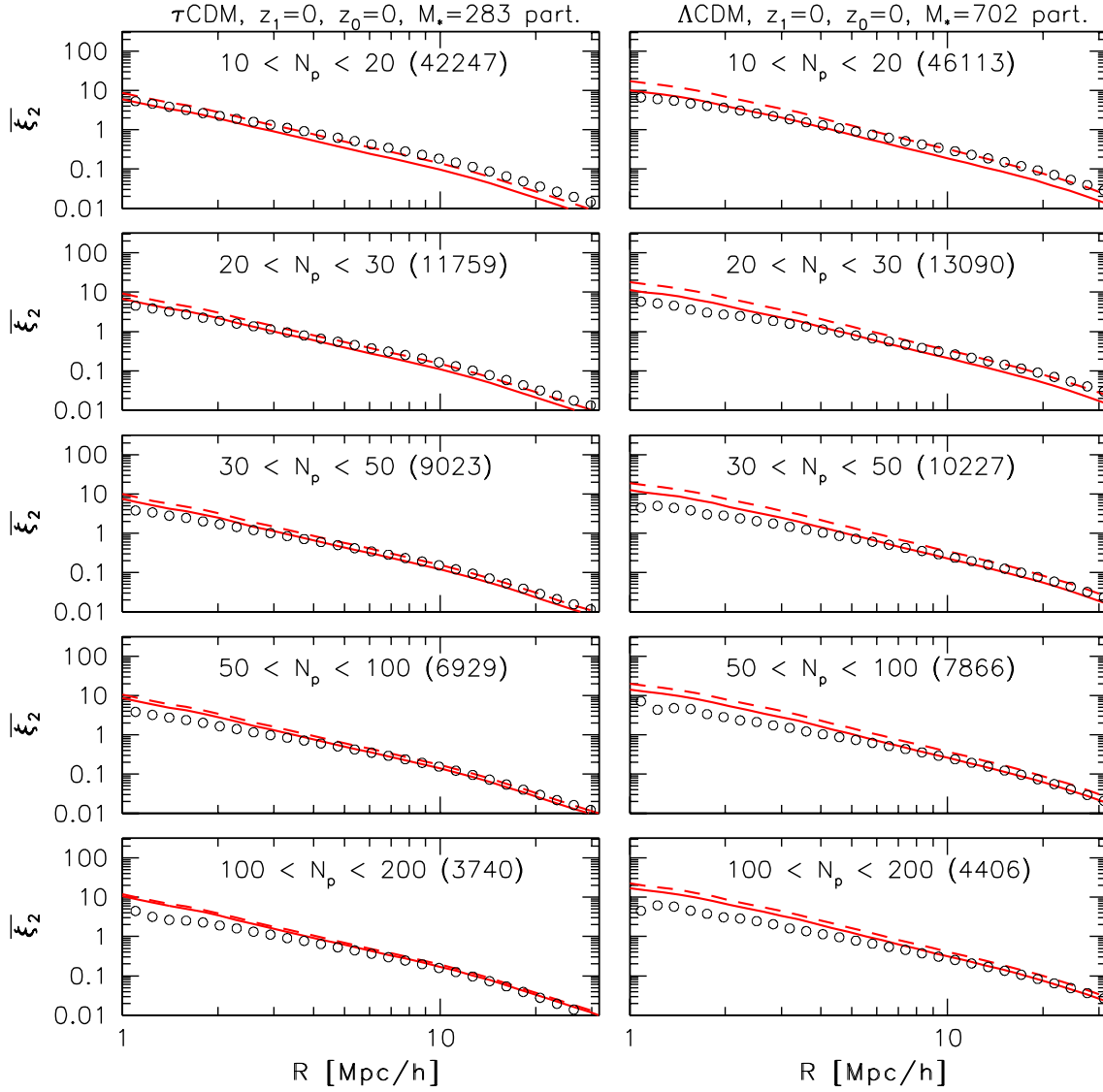
Figure 2.4: Variance $\overline{\xi}_2$ of dark haloes with different mass ranges obtained from the counts-in-cells analysis (symbols) and from applying the bias model from MW (solid line) and its SMT extension (dashed-line). The variance of the mass obtained directly from the simulations is shown by the dotted line. Results are shown for the GIF ΛCDM model and for haloes identified and analyzed at the present time.

Figure 2.5: Variance $\overline{\xi}_2$ of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MW (solid line) and its SMT extension (dashed-line). The variance of the mass obtained directly from the simulations is shown by the dotted line. Results are shown for the GIF ΛCDM model and for haloes identified at $z = 1$ and analyzed at the present time.

Figure 2.6: Variance $\overline{\xi}_2$ of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MW (solid line) and its SMT extension (dashed-line). The variance of the mass obtained directly from the simulations is shown by the dotted line. Results are shown for the ΛCDM model and for haloes identified at $z = 3$ and analyzed at the present time.

Figure 2.7: Variance $\overline{\xi}_2$ obtained from the counts-in-cells analysis (symbols) and from applying the bias model from MW (solid line) and its ellipsoidal collapse extension (dashed-line) of haloes less massive than M_*. Each row in the panel corresponds to a different range of halo masses, as indicated in the boxes.

2.4.3 Testing the models for the Higher-Order Moments

Following the procedure given in section 2.2 we have obtained the volume-averaged correlation functions up to the fourth order from the mass distribution and from the various halo samples. Analyses have been performed for two different cases. In the first, the high-order moments are calculated at the same time when the dark haloes are identified. In the second, haloes are identified at some high redshift while the calculations of the high-order moments are performed for their descendants at a later time. In all cases, the redshift at which halo identification is made is denoted by z_1, while the redshift at which the high-order moments are calculated is denoted by z_0.

Figures (2.8)–(2.11) show the third- and fourth- order moments from the VIRGO ΛCDM simulation, together with model predictions. Both the MJW model and the SMT extension work remarkably well, especially in the two epoch case (where $z_1 > z_0$). The difference between the predictions of the MJW model and the SMT extension is not large for the VIRGO simulation, because VIRGO haloes are quite massive due to the relatively low mass resolution. In these two figures we also plot the prediction of the MJW model with S_3 and S_4 given by the perturbation theory (see Bernardeau 1994). The fact that this prediction also matches the simulation results suggests that the moments obtained from the VIRGO simulations are not affected significantly by the finite-volume effect and that the MJW model is a good approximation to the high-order moments for haloes that are not much smaller than $M_\star$ [defined by $\sigma(M_\star) = 1.68$]. Similarly, Figures (2.4)–(2.6) show the variance from the GIF ΛCDM simulation, together with the predictions from the MW and its ellipsoidal collapse extension.

As already stated in the last section, the GIF simulations allow one to test the theoretical models for haloes with mass $M \ll M_\star$. Since the GIF simulations have relatively small simulation boxes, the high-order moments are expected to be affected significantly by the finite-volume effect (Colombi et al. 1994). However, this effect in each simulation is expected to be similar for both the mass distribution and the halo distribution. Thus, to test the bias model given in (2.37) and (2.38) by a numerical simulation we should use the value of S_3 and S_4 obtained directly from the simulation, because it is the simulated power spectrum (not the theoretical spectrum) that is responsible for the clustering in the simulation. Figures (2.16)–(2.17) show the results obtained for the GIF simulations. As one can see, there is a good agreement between model predictions and simulation

results. For haloes with masses much smaller than $M_\star$, the MJW model underestimates the skewness and kurtosis, while the SMT extension gives a much better fit (Figures 2.16 and 2.17). Thus, the SMT extension not only improves the models for the mass function and second-order moment of dark haloes, but also improves the models for the high-order moments. This gives further support to the notion that the ellipsoidal model is a better approximation to the formation of dark haloes in the cosmological density field than the spherical model.

From a comparison between the VIRGO and GIF results, it is evident that both the skewness and kurtosis are strongly affected by the finite-volume effect. However, if the loss of clustering power due to the finite volume is taken into account, the model predictions are in good agreement with the numerical results, suggesting that the bias relations given by (2.37) and (2.38), with the coefficients given by the extended Press-Schechter formalism, are good approximations to the skewness and kurtosis of dark haloes in the quasi-linear regime.

To see more clearly the difference between the MJW model and the SMT extension, we show in figure 2.18 the amplitudes of the halo skewness and kurtosis at a fixed radius ($R = 10h^{-1}\mathrm{Mpc}$) as a function of the linear bias parameter $b = b_1$ [see equations (2.42) and (2.47)]. The curves correspond to the predictions from the models for the present-day descendants of haloes at three values of z_1 (3.0, 1.0 and 0.0). From the figure we see that in all cases the values of $S_{j,h}$ are lower than those for the mass unless b is comparable to or smaller than 1. This result was obtained in MJW based on the spherical model. We see that this is also true even if the SMT extension is used, although the amplitudes of $S_{j,h}$ given by the elliptical model are higher than those given by the spherical model for a given b. These features in $S_{3,h}$ and $S_{4,h}$ have been used in MJW to constrain the bias parameter b for galaxies.
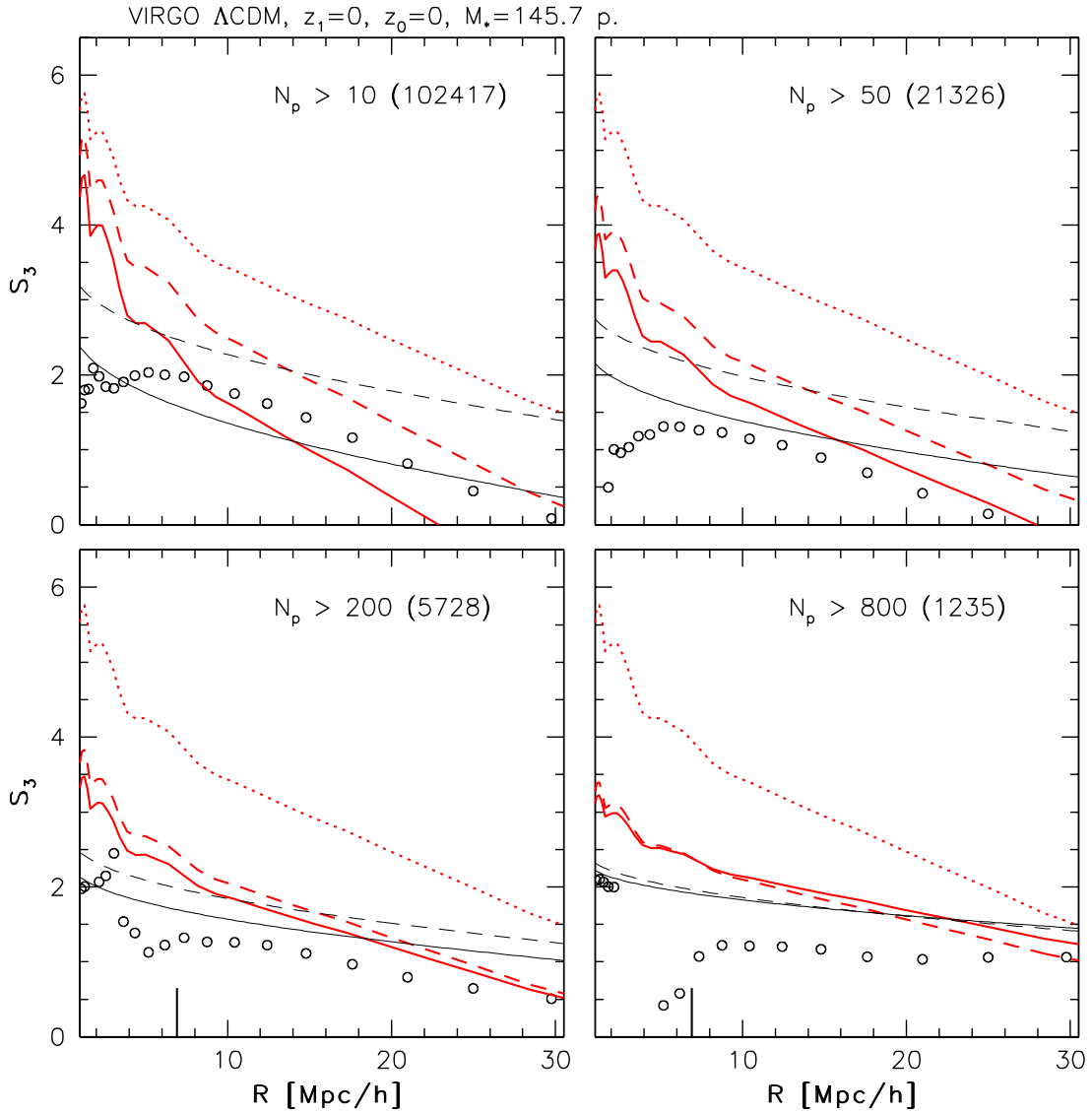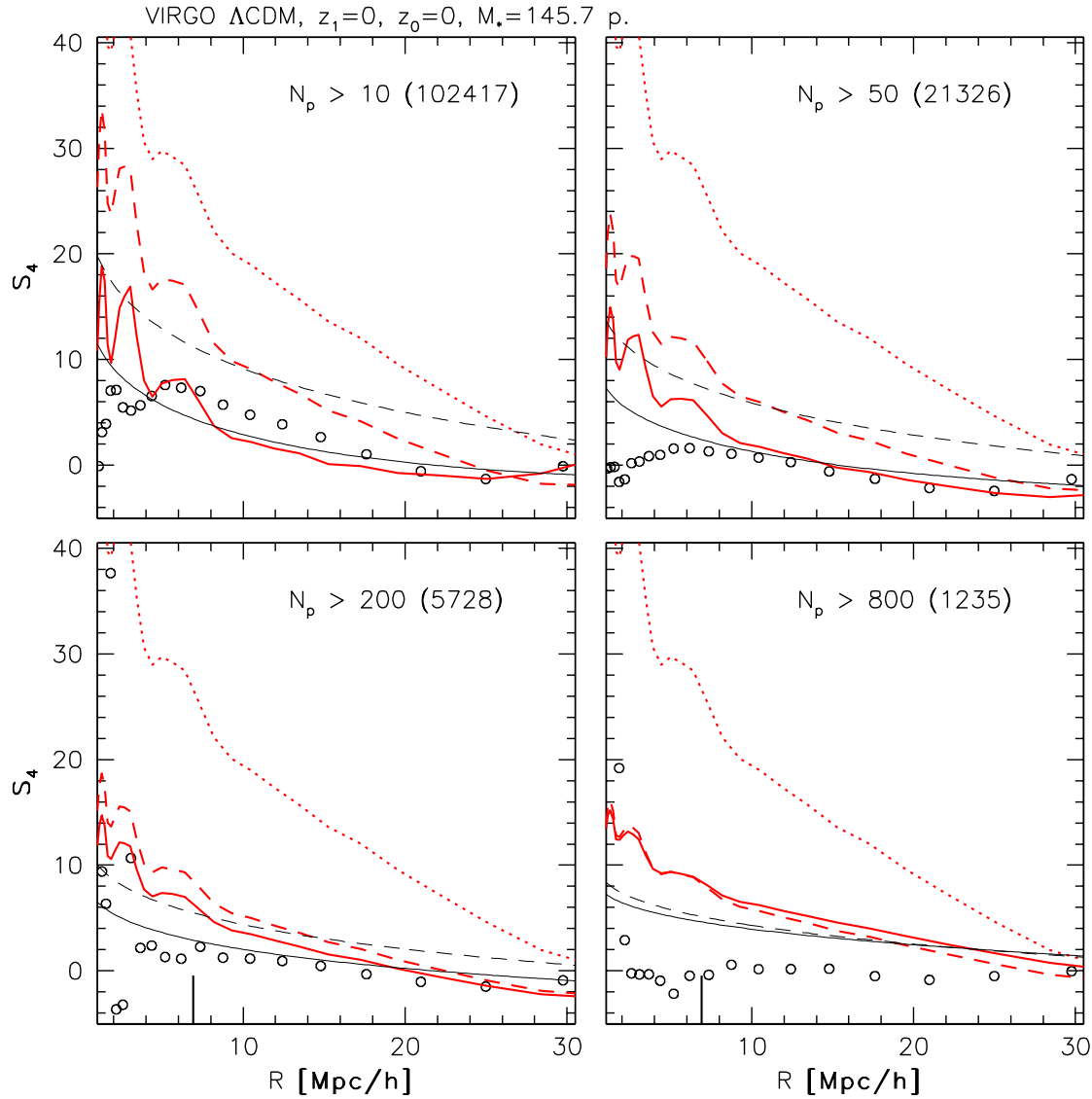
Figure 2.8: Skewness S_3 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols), from applying the bias model from MJW (solid line) and its SMT extension (dashed-line). The moments for the mass distribution are shown by the dotted line and the moments for the haloes obtained using the moments for the mass from the perturbation theory are shown as a dot-long dashed line. The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the VIRGO ΛCDM simulations. The haloes have been identified and analyzed at the times written in the upper-left boxes. The value of M_* is also included. Each box corresponds to a different range of masses of haloes, as appearing in the labels. The quantities between parenthesis correspond to the number of haloes in each sample.
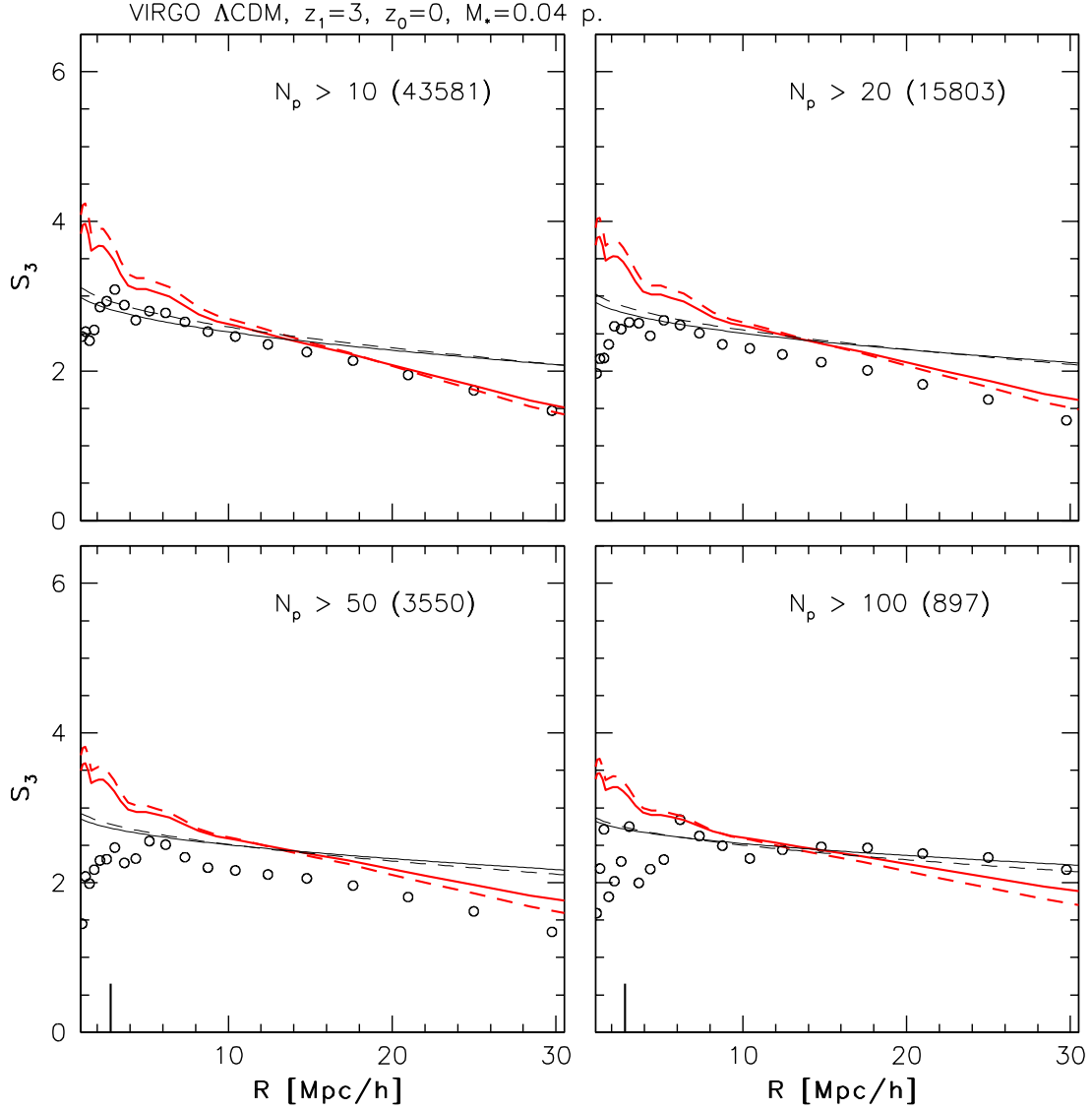
47

Figure 2.9: Kurtosis S_4 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols), from applying the bias model from MJW (solid line) and its SMT extension (dashed-line). The moments for the mass distribution are shown by the dotted line and the moments for the haloes obtained using the moments for the mass from the perturbation theory are shown as a dot-long dashed line. The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the VIRGO ΛCDM simulations. The haloes have been identified and analyzed at the times written in the upper-left boxes. The value of M_* is also included. Each box corresponds to a different range of masses of haloes, as appearing in the labels. The quantities between parenthesis correspond to the number of haloes in each sample.

Figure 2.10: Skewness S_3 for haloes in the VIRGO ΛCDM simulations for haloes identified at $z = 3$ and analyzed at the present time. The lines, symbols correspond to the same models and quantities as in figure 2.8.
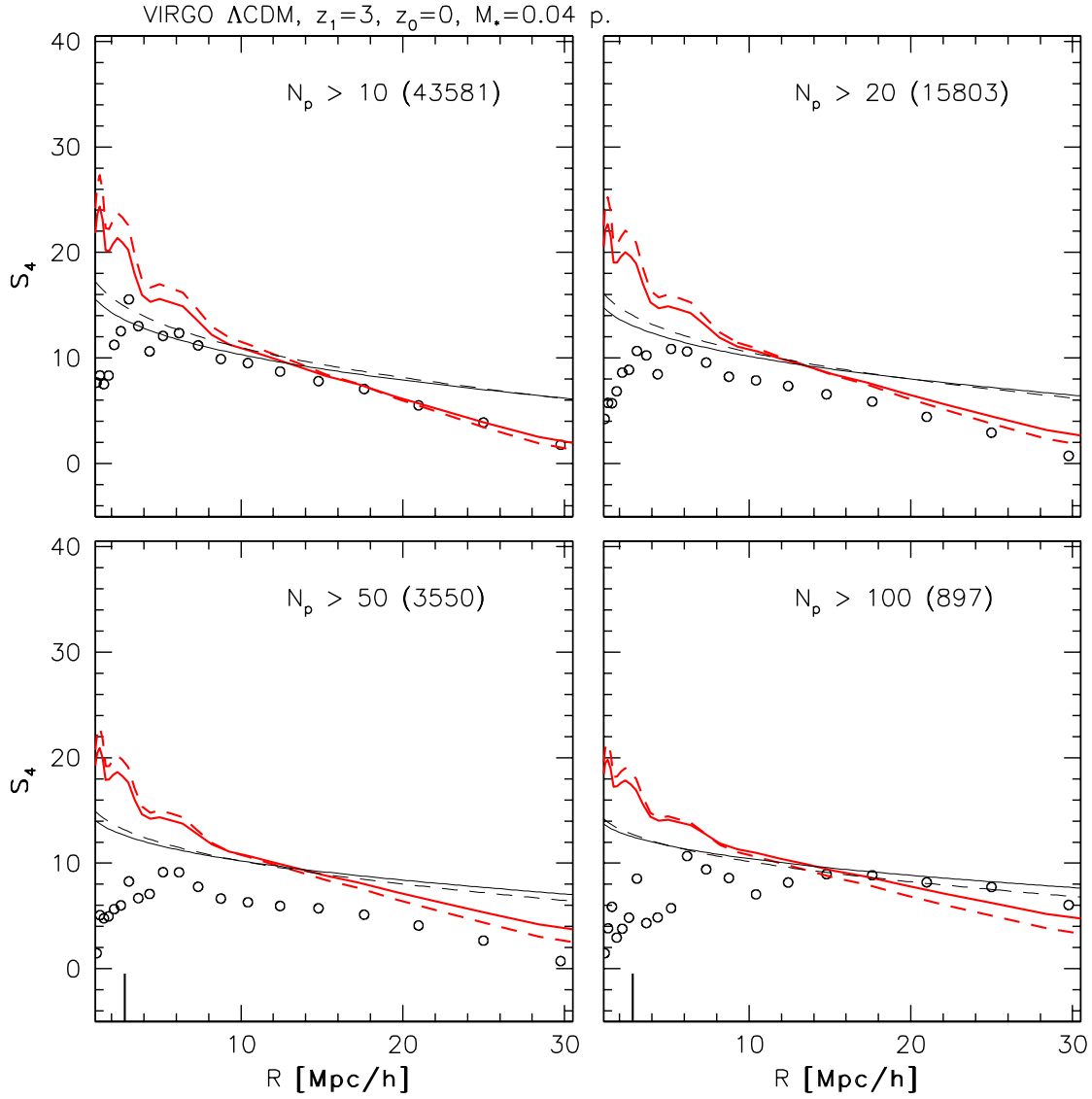
Figure 2.11: Kurtosis S_4 for haloes in the VIRGO ΛCDM simulations for haloes identified at $z = 3$ and analyzed at the present time. The notation is the same as in figure 2.9.
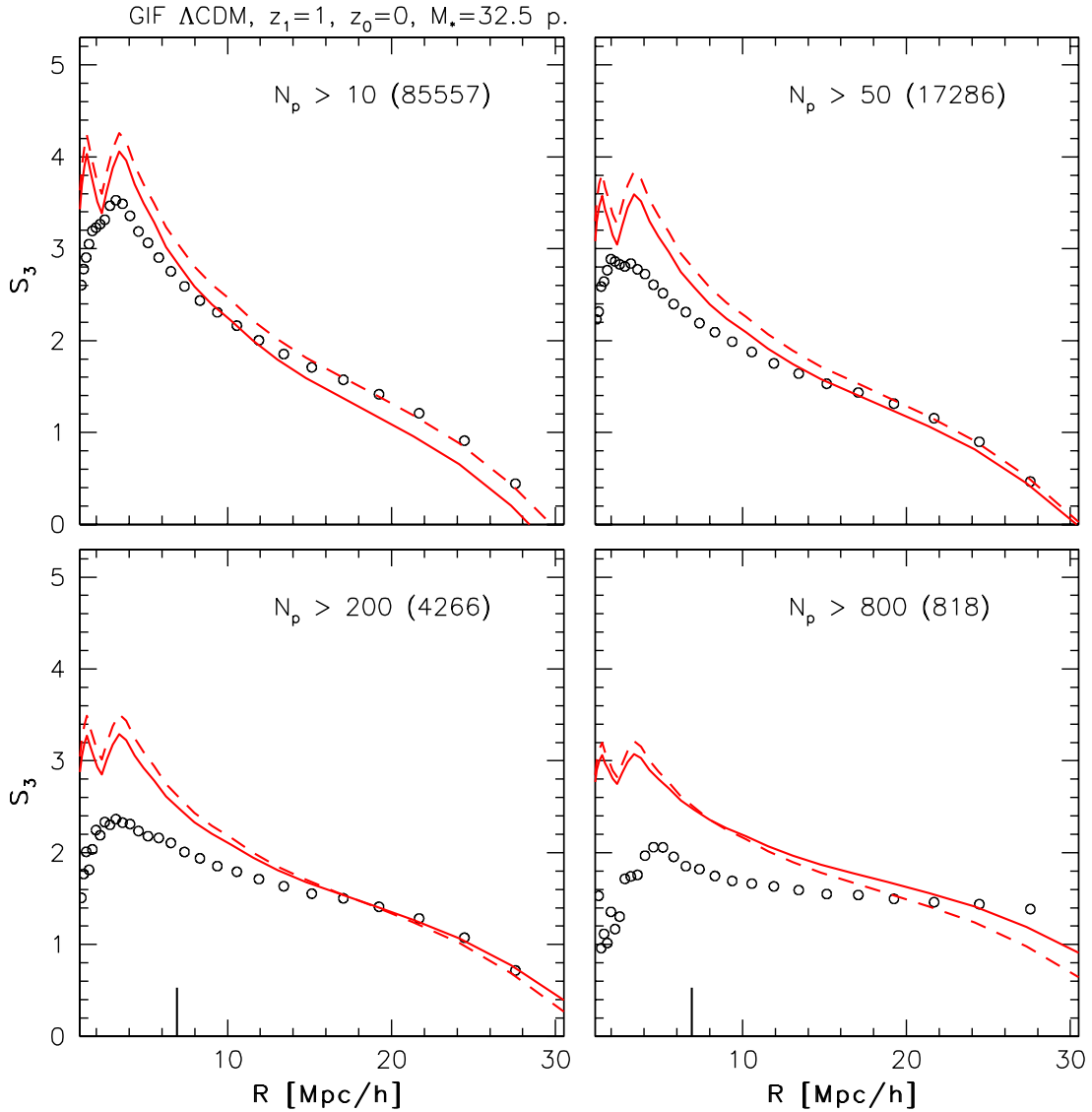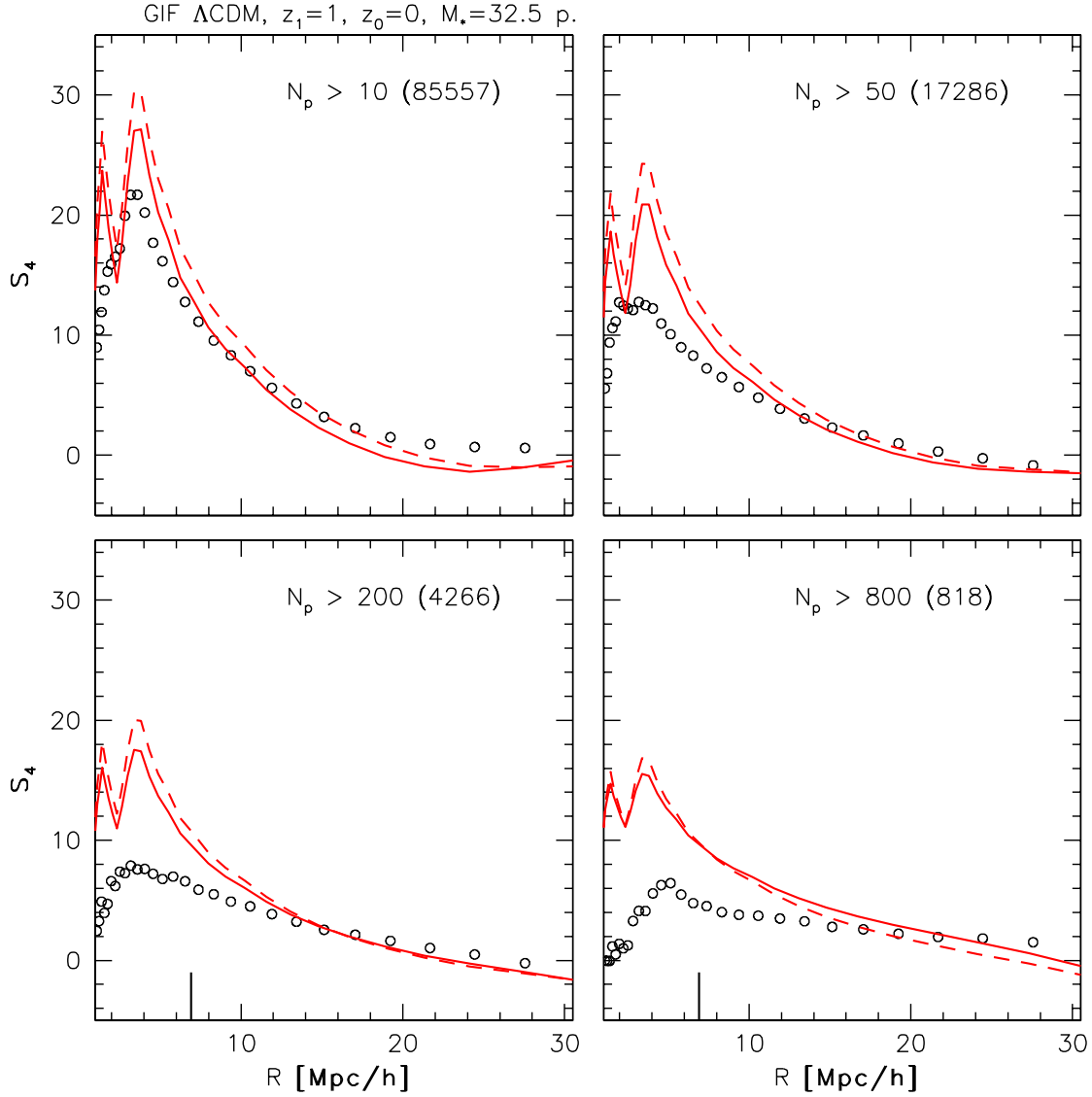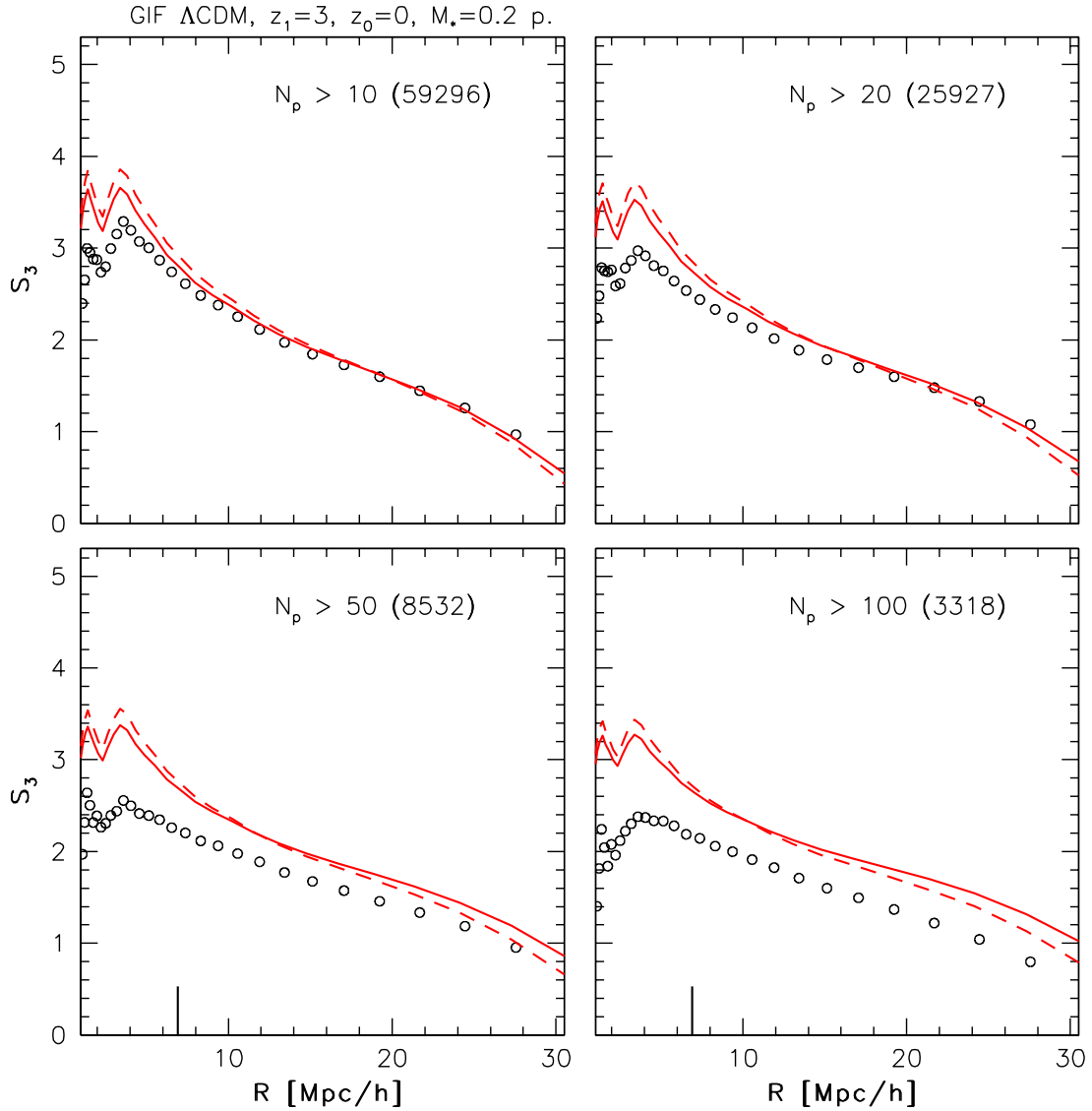
Figure 2.12: Skewness S_3 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line). The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the GIF ΛCDM model and for haloes identified at $z = 1$ and analyzed at the present time.
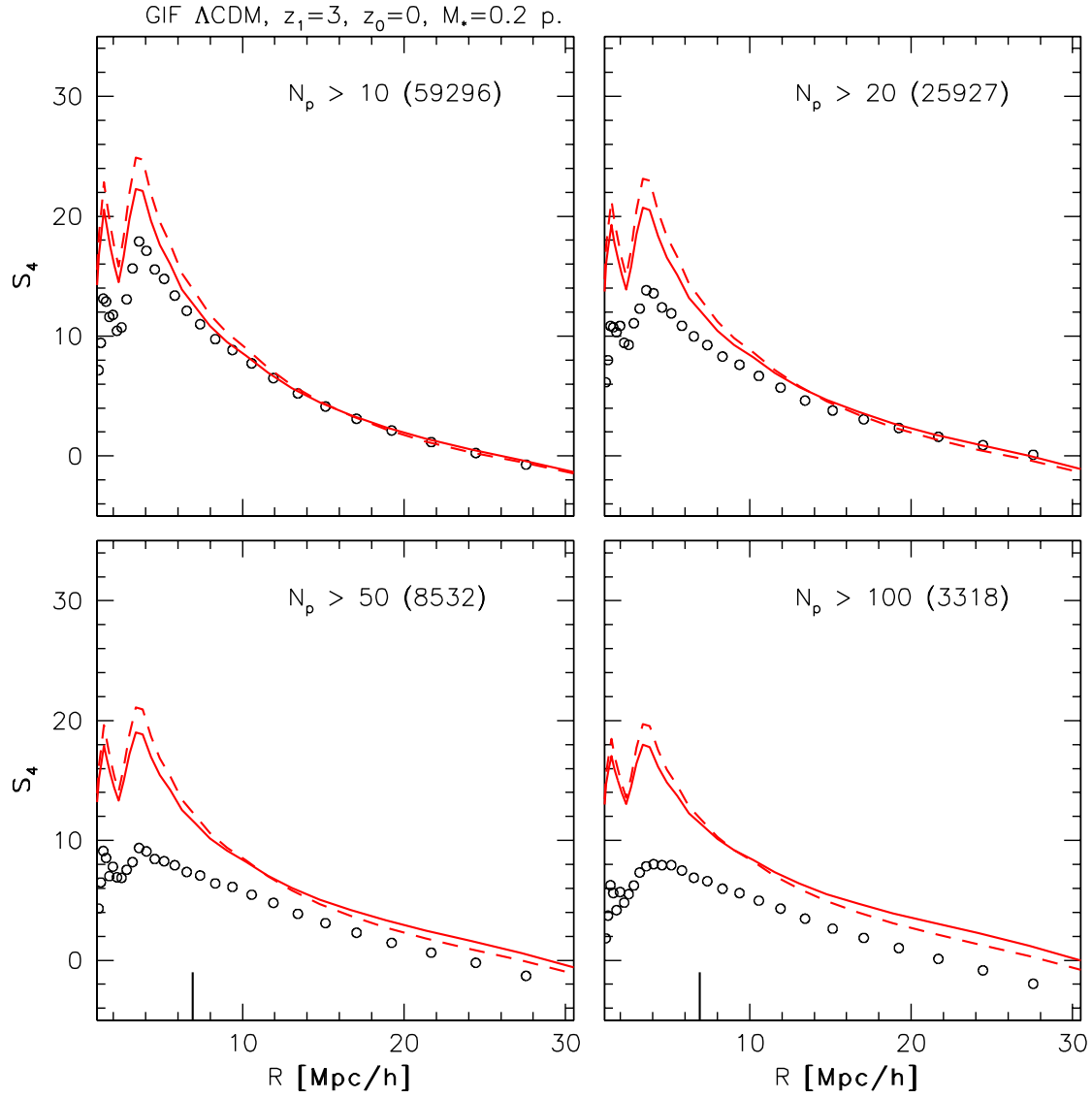
Figure 2.13: Kurtosis S_4 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line). The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the GIF ΛCDM model and for haloes identified at $z = 1$ and analyzed at the present time.

Figure 2.14: Skewness S_3 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line).The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the ΛCDM model and for haloes identified at $z = 3$ and analyzed at the present time.
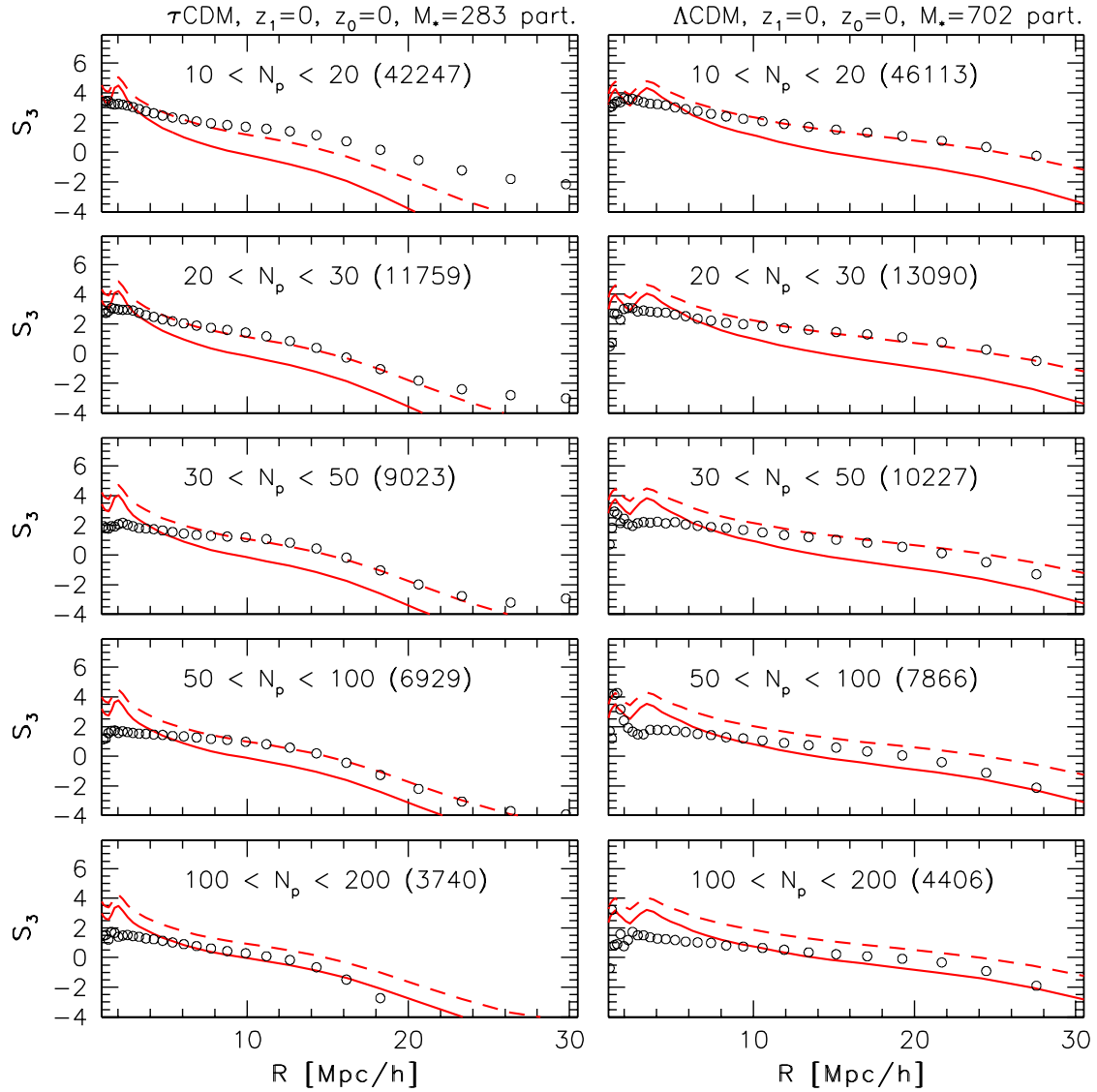
Figure 2.15: Kurtosis S_4 of dark haloes with different mass ranges obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line).The thick ticks on the horizontal axis show the scales where $\overline{\xi}_2(R) = 1$. Results are shown for the ΛCDM model and for haloes identified at $z = 3$ and analyzed at the present time.

Figure 2.16: Skewness S_3 obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line) of haloes less massive than M_*. Each row in the panel corresponds to a different range of halo masses, as indicated in the boxes.
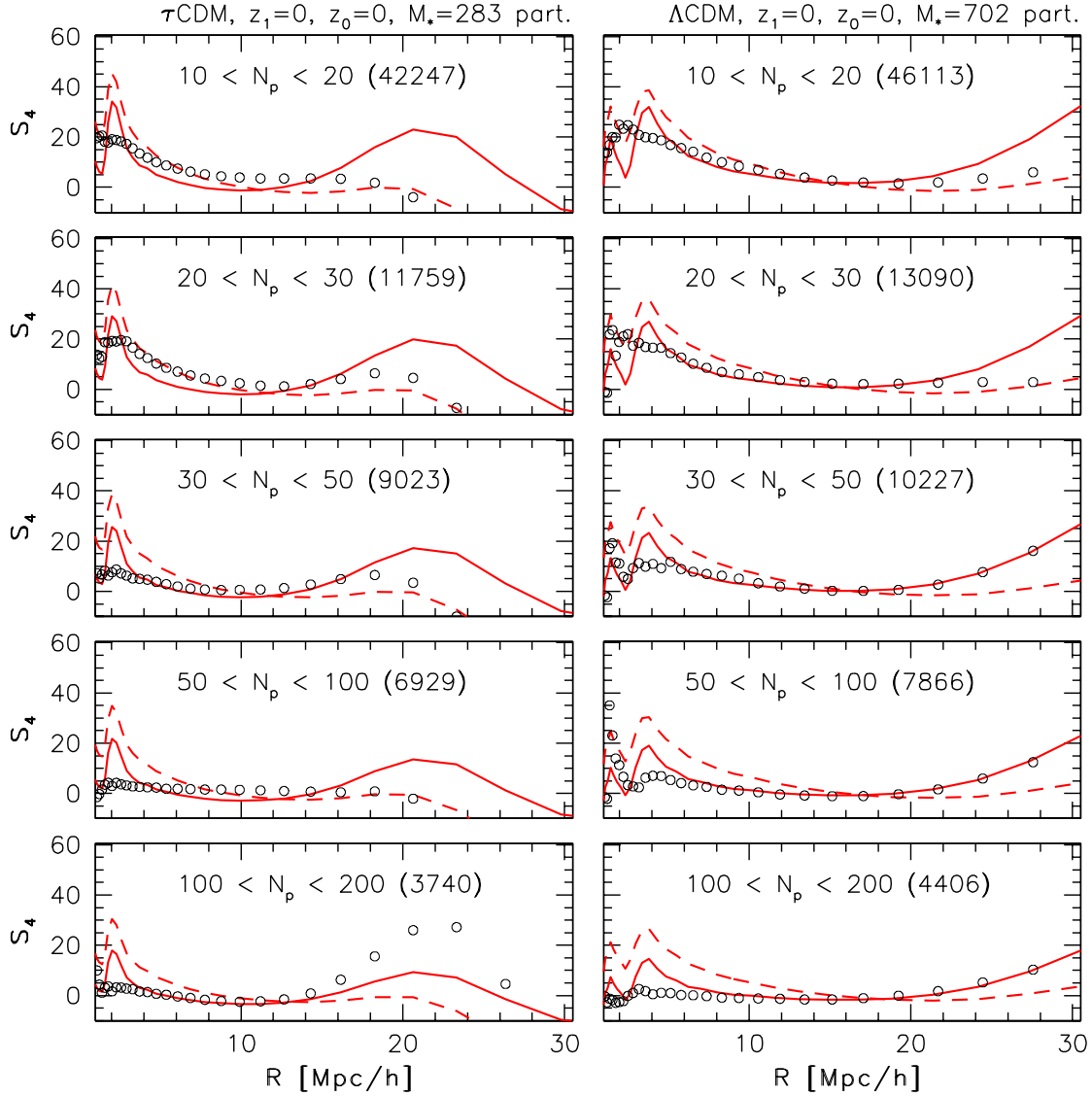
Figure 2.17: Kurtosis S_4 obtained from counts-in-cells analysis (symbols) and from applying the bias model from MJW (solid line) and its SMT extension (dashed-line) of haloes less massive than M_*. Each row in the panel corresponds to a different range of halo masses, as indicated in the boxes.
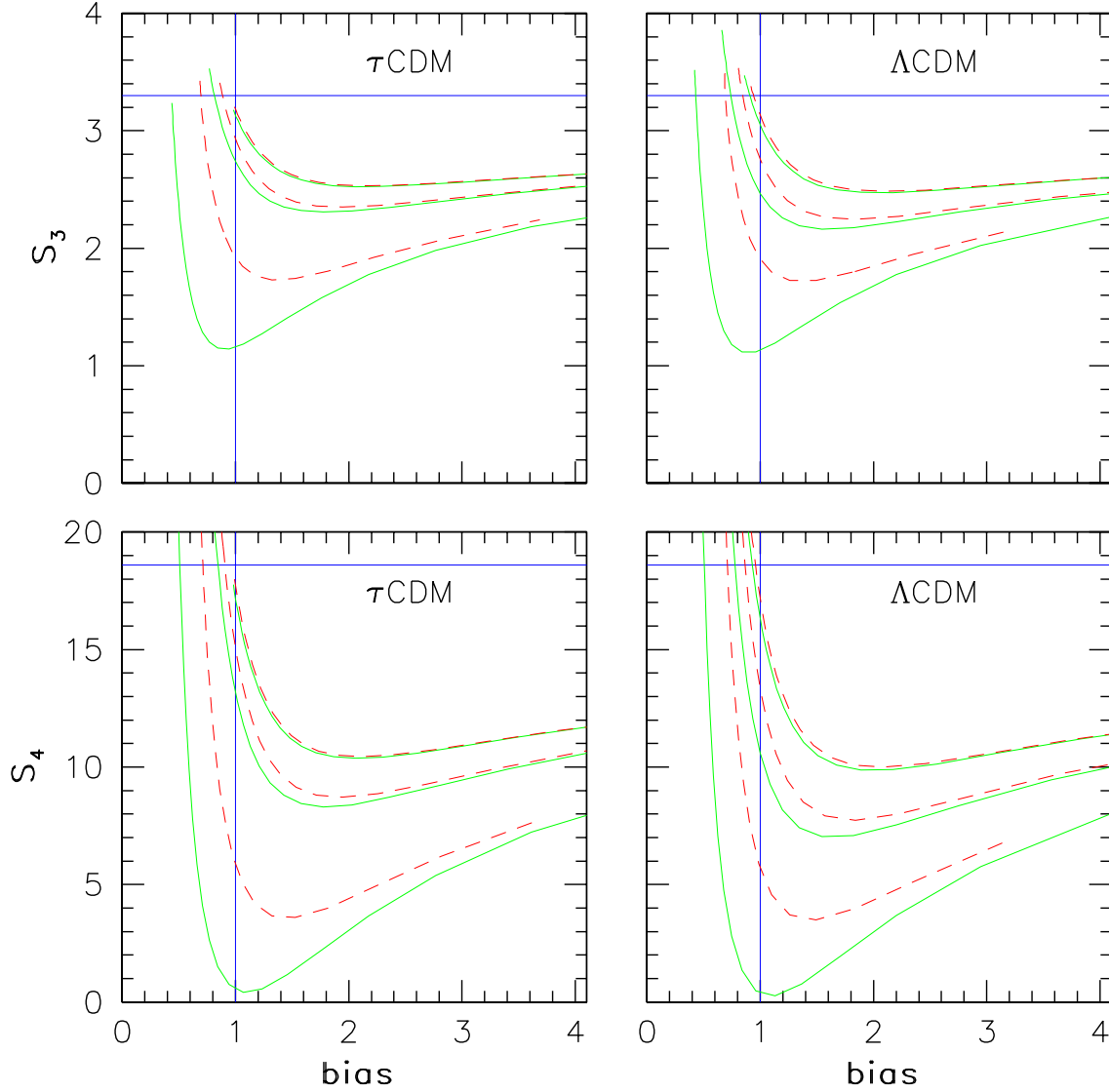
Figure 2.18: Predictions from the MJW model (solid lines) and its SMT extension (dashed-line) for the skewness and kurtosis of haloes at a radius $R = 10\ h^{-1}\ Mpc$ as a function of the linear bias parameter b. Each pair of curves shows the results for a given δ_1, where $z_1 \equiv (\delta_1/1.686 - 1) = 0.,\ 1.0,\ 3.0$) from bottom to top.

2.5   Discussion

From the results shown in figure 2.18 we see that for present time descendants of haloes already formed at a given redshift ($z > 0$), the values of the skewness and kurtosis depend only weakly on the object mass if the bias parameter b is larger than or near to one . On the other hand, in the same range of b it is clear that the high order moments depend on the identification redshift, which is associated to the redshift of formation of the objects, with the corresponding values increasing as the formation redshift increases. Therefore the values of S_3 and S_4 of old objects, like elliptical galaxies, are expected to be higher than the corresponding moments of more recently formed objects, such as spiral galaxies. This feature can be useful in studying different galaxy populations.

We have used the models to analyze the predicted values of the high order moments for high redshift objects, like the Lyman Break Galaxies (LBG), which are commonly assumed to form in the center of the most massive haloes at redshift ~ 3 (Mo & Fukugita 1996; Adelberger et al. 1998; Jing & Suto 1998; Mo et al. 1999). Under this assumption and, supposing that only a negligible fraction of those haloes host a secondary observable galaxy the observed LBGs correspond to the most massive haloes at $z \sim 3$. We have estimated the predicted values for the skewness and kurtosis at a fixed scale $R = 10 \ h^{-1} \ Mpc$ of the LBGs ($z = 3$) and their descendants at a given redshift z. We chose this value of R, because the mass density in the universe is still in the quasi-linear regime and the high order moments of galaxy distributions are more difficult to measure on much larger scales.

For our estimates we have used the coefficients as given by equations (2.42)-(2.44), where the S_q (q = 3,4) for the mass distribution are obtained from linear perturbation theory (Bernardeau 1994). and the weighted average needed to get the effective b_k's is done by means of the mass function from the Press-Schechter formalism. The main parameter for the estimation of the b_k's for the LBGs corresponds to the observed abundance of LBGs, namely the number density given by (Adelberger et al. 1998). This number is $N_{lbg} \approx 8 \times 10^{-3} h^3 Mpc^{-3}$ at $z \sim 3$ for an Einstein-de Sitter universe, and is similar to the present abundance of L_* galaxies. The corresponding number for the ΛCDM universe is estimated by multiplying this number by the comoving volume per unit redshift at $z \sim 3$ for an Einstein-de Sitter universe divided by the corresponding value for the ΛCDM universe.

In figure (2.19) we show the values of the skewness, kurtosis and linear bias at $R = 10\ h^{-1}\ Mpc$ of the LBGs, as a function of the redshift, in the ΛCDM and τCDM models. From the curves we see that, although the linear bias parameter is quite different in both CDM models, the values obtained for the moments are too similar to be used as a tool to constrain cosmological parameters.

2.6   Summary

The spherical collapse based models for the moments of dark matter halo counts-in-cells from Mo & White (1996) (variance) and Mo et al. (1997) (higher-order moments), as well as their ellipsoidal collapse extension by Sheth et al. (2001) have been tested using two sets of high-resolution N-body simulations with different simulation boxes and mass resolution. From the set with very large simulation boxes, which allows us to control the finite volume effect, it has been found that the models work remarkably well for CDM universes. The good performance of the models when the moments from the mass distribution are estimated using the linear perturbation theory, shows that the moments from this set (VIRGO Simulations) are practically unaffected by the finite volume effect. The other set of simulations, having much higher mass resolution, has been used to test the models for low-mass haloes, showing that significant improvement can be achieved for haloes less massive than $M^{\star}$ if the ellipsoidal collapse model is used instead of the spherical collapse model in defining dark haloes and that for massive haloes both the MW and MJW models and their ellipsoidal extension work remarkably well.

The theoretical model has been used to predict the high-order moments at a fixed scale of the Lyman break galaxies observed at $z = 3$ and their descendants at lower redshifts. It has been found that, although the linear bias parameter b depends strongly on the cosmology adopted, the values of the high-order moments are practically the same in both CDM models, and therefore the high-order moments from the spatial distribution of these objects cannot be used to constrain cosmological parameters.
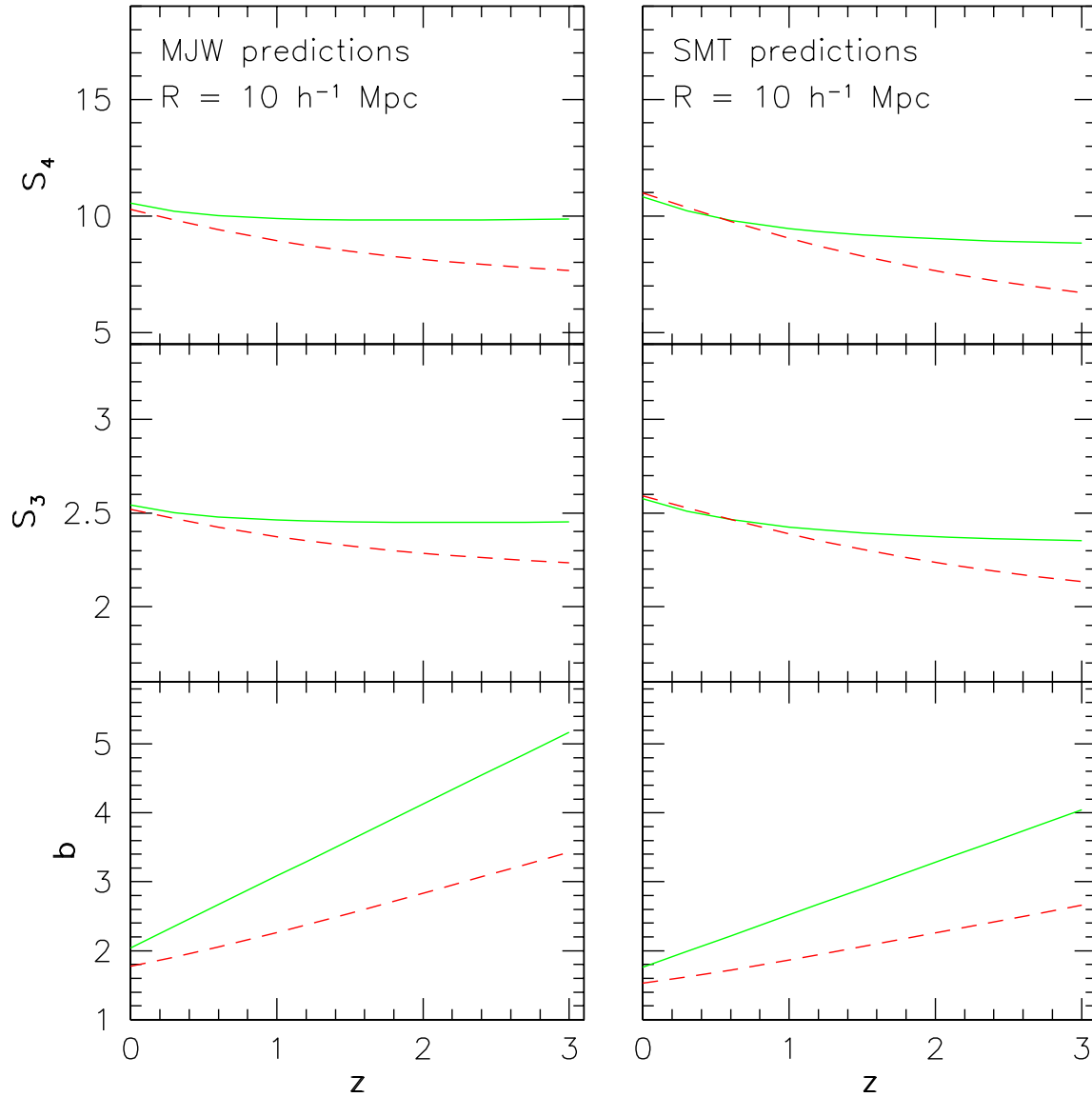
Figure 2.19: Skewness and kurtosis at $R = 10h^{-1}$ Mpc for the LBGs at $z = 3$ and their descendants at later epochs. The curves correspond to the ΛCDM model (dashed lines) and the τCDM model (solid lines). The left panel shows the predictions from the MJW model and the right panel shows the predictions from the SMT extension. The horizontal line shows the corresponding value for S_3 or S_4 of the mass at $z = 0$.

Chapter 3

Stochastic Bias

3.1 Introduction

In chapter 2 I have discussed the effect of a deterministic bias for the halo and galaxy distribution. In this chapter I want to discuss the effect of a stochastic bias relation.

Stochasticity may be important in high-order statistics as well as in the full distribution function of haloes. In fact, the stochastic nature of the bias relation was already emphasized in the original paper of MW; in particular, MW pointed out that halo-exclusion can cause sub-Poisson variance. Sheth & Lemson (1999) showed how the effects of stochasticity could be incorporated, easily and efficiently, into the analysis of the higher order moments.

Recently Somerville et al. (2001) used N-body simulations to study the stochasticity and non-linearity of the bias relation based on the formalism developed by Dekel & Lahav (1999). They analyzed the bias relation for haloes with masses larger than $1.0 \times 10^{12} \, h^{-1} \, M_\odot$ in spherical volumes of radius $8 \, h^{-1}$Mpc. The present work is quite closely related to theirs but contains several distinct aspects. First of all, this analysis is focused on the distribution function $P_V(N|\delta_m)$, which gives the probability of finding N haloes in a volume V with mass density contrast δ_m [$\delta_m \equiv \frac{\rho}{\bar{\rho}} - 1$, where ρ is the mass density and $\bar{\rho}$ is the mean mass density]. As it will be shown later, this function completely specifies the relation between the spatial distribution of haloes and that of the mass in a statistical sense. Second, the present analysis covers a wider range of halo masses and a larger range of volumes for the counts-in-cells. Finally, an attempt to develop a theoretical model to describe the

stochasticity of the bias relation is performed. This theoretical model is based on the mean bias relation given in MW and on the variance model given in Sheth & Lemson (1999). As it will be seen below, the original Sheth & Lemson model fails in high mass density regions, where gravitational clustering becomes important.

3.2 The Halo-Mass Bias Relation

3.2.1 The Conditional Probability Function

Dark matter haloes are formed in the cosmological density field due to nonlinear gravitational collapse. In general, the halo density field is expected to be correlated with the underlying mass density field. Thus, if we denote by δ_m the matter density fluctuations field and by N the halo number (where both fields are smoothed in regions of some given volume), N and δ_m are related. We refer to this relation as the halo bias relation, because it describes how the halo distribution is biased with respect to the underlying mass distribution. Since in general the halo number in a volume depends not only on the mean mass density but also on other properties (such as the clumpiness) of the mass distribution, the relation between N and δ_m is not expected to be deterministic. It must be stochastic. The stochasticity of the bias relation can be described by the conditional distribution function, $P_V(N|\delta_m)$, which gives the probability of finding N haloes in a volume V with mass density contrast δ_m. This conditional probability completely specifies the relation between the mass and halo density fields in a statistical sense. Indeed, once $P_V(N|\delta_m)$ is known, the full count-in-cell function $P_V(N)$ for haloes can be obtained from the mass distribution function $P_V(\delta_m)$ through

$$P_V(N) = \int_{-\infty}^{\infty} P_V(N \mid \delta_m) P_V(\delta_m) \, d\delta_m. \tag{3.1}$$

The form of $P_V(N|\delta_m)$ depends on how dark haloes form in the cosmological density field and is not known *a priori*. The simplest assumption is that it is Poissonian. This assumption is in fact used in almost all interpretations of the moments of galaxy counts in cells (c.f. Peebles 1980), where terms of Poisson shot noise are subtracted to obtain the correlation strength of the underlying density field. However, this assumption is not solidly based, and so it is important to examine if other assumptions on the form of

$P_V(N|\delta_m)$ actually work better for dark haloes.

Some Analytical Functions

For the present analysis I have chosen four simple possible analytical forms for the conditional probability. These functions are fitted to the conditional probability measured from the simulations, in order to investigate which of them describes better the numerical halo-mass bias relation. The standard functions, chosen for this investigation, correspond to the Poisson, Gaussian, Lognormal and the Thermodynamic Distribution functions (Saslaw & Hamilton 1984). The last one is derived from thermodynamic foundations and has the property that it converges to a Poisson distribution when the b_{Td} parameter is zero (see equation (3.5)). These functions are expressed in the following way:

- Poisson Distribution

$$P_P(N \mid \delta) = \frac{\lambda^N \, e^{-\lambda}}{N!}. \tag{3.2}$$

- Normal Distribution

$$P_N(N \mid \delta) = \frac{1}{\sqrt{2\pi} \, \sigma} \, \exp\left[-\frac{(N-\lambda)^2}{2\sigma^2}\right], \tag{3.3}$$

  where σ corresponds to the variance of the distribution.

- Lognormal Distribution

$$P_{LN}(N \mid \delta) = \frac{1}{\sqrt{2\pi} \, \sigma \, N} \, \exp\left[-\frac{(\ln N - \lambda)^2}{2\sigma^2}\right], \tag{3.4}$$

  where σ corresponds to the variance of the underlying Normal distribution and $N > 0$.

- Thermodynamic Distribution

$$P_{Td}(N \mid \delta) = \frac{\lambda(1 - b_{Td})}{N!}(\lambda(1 - b_{Td}) + Nb_{Td})^{N-1} \, \exp\left[-\lambda(1 - b_{Td}) - Nb_{Td}\right], \tag{3.5}$$

  where $0 \leq b_{Td} < 1$ is a free parameter. Notice that for $b_{Td} = 0$ this distribution corresponds to a Poissonian one.

In principle, the mean of the bias relation is fixed by the fact that the halo distribution function obtained from the right-hand-side integral in equation (3.1) must reproduce the actual mean number of haloes in the volume. However, for the fitting process the parameters of the corresponding distributions are used as purely free fitting parameters.

3.2.2   A Model for the Halo-Mass Bias Relation

To second order, the probability distribution function $P_V(N|\delta_m)$ is described by the mean bias relation $N = N(\delta_m)$ and the variance $\sigma^2 \equiv \langle N^2|\delta_m \rangle$.

The Mean Bias Relation

Mo & White (1996) developed a model for the mean bias relation of haloes based on the spherical collapse model. Their model works well for massive haloes and an extension of it by Sheth et al. (2001) based on ellipsoidal collapse may work better for low mass haloes. The model has been already introduced in section (2.3.1). Briefly, the mean of the bias relation from the MW model is given by

$$\delta_h(1 \mid 0) = \left(1 + \frac{\nu_1^2 - 1}{\delta_1} \right) \delta. \tag{3.6}$$

The Variance of the Bias Relation

Sheth & Lemson (1999) have presented a model for the variance of the bias relation which accounts for the halo exclusion due to the finite size of haloes (i.e. two different haloes can not occupy the same volume). They have shown that their model is able to describe the first and second moments of the halo distribution from scale-free N-body simulations. Nevertheless the model is expected to fail when the underlying clustering makes a significant contribution to the variance. As an amendment, an additional term accounting for the clustering of haloes in high density regions is introduced.

Briefly, in the initial Lagrangian space each halo occupies a volume proportional to its mass, and haloes can not overlap. This fact implies that the halo distribution is affected by volume exclusion effects, especially at scales smaller than the typical size of a halo. The model by Sheth & Lemson (1999) allows one to include explicitly the exclusion effects at computing the moments of the Lagrangian space halo distribution. They combined their model with the Mo & White (1996) spherical collapse model to quantify the dynamical evolution of the mean and variance of the bias relation. In their paper Sheth & Lemson (1999) presented a detailed treatment of clustering from Poisson and white-noise Gaussian initial condition, since these cases have exact analytical results and conjectured that these results can be easily extended to obtain an accurate approximated model for the bias associated with the clustering from more general Gaussian initial density fields.

Briefly, the volume exclusion effect in the model is introduced as follows. Let us suppose that there are n haloes with mass M_1 within a spherical region with mass M_0.
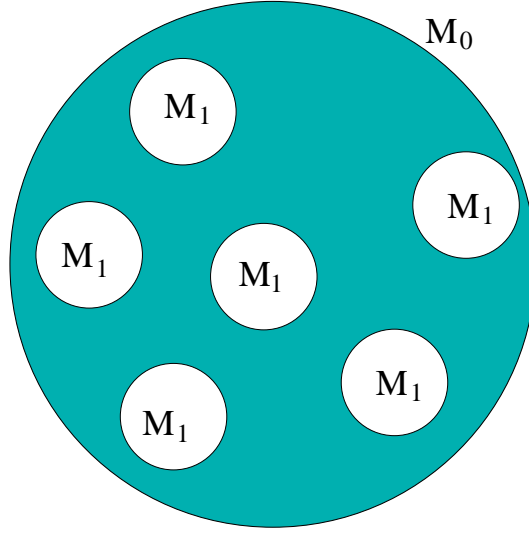

Figure 3.1: 2–D Scheme of the volume occupancy of n M_1 haloes in a spherical region of mass M_0

The average overdensity of the whole volume is

$$1 + \delta_0 = \frac{M_0}{\overline{\rho} V_0} \tag{3.7}$$

and the average overdensity of the remaining volume [shaded area in figure (3.1)] is

$$1 + \delta^{(n)} = \frac{M_0 - nM_1}{\overline{\rho}(V_0 - nV_1)} \quad \text{and} \quad \delta^{(0)} \equiv \delta_0. \tag{3.8}$$

From the definition of the mass overdensity δ, one has that the mass contained within the filter of volume V can be written

$$M \equiv \overline{\rho} V_1 (1 + \delta). \tag{3.9}$$

Notice that it is assumed that $\Delta \ll 1$. Thus, there is no problem with this definition of the mass within the filter because $\|\delta\| \ll 1$ almost surely and the possibility that $\delta < -1$ is extremely unlikely.

Since $M_1 \equiv \overline{\rho} V_1 (1 + \delta_1)$ it can be shown that

$$\delta^1 - \delta^{(n)} = (\delta_1 - \delta_0) \frac{M_0}{M_0 - nM_1}, \tag{3.10}$$

to the lowest order in the δ–terms.

With this definition one has the expression for the i–th factorial moment

$$\phi_i(M_1, \delta_1 \mid M_0, \delta_0) = \prod_{n=0}^{i-1} N(M_1, \delta_1 \mid M_0 - nM_1, \delta^{(n)}), \qquad (3.11)$$

for $iM_1 \leq M_0$ and zero otherwise.

Sheth & Lemson (1999) define further the i–th factorial moment of the corresponding halo counts-in-cells distribution (Ξ_i) by

$$\int_{-\infty}^{\delta_1} \phi_i(M_1, \delta_1 \mid M_0, \delta_0) \; q(\delta_1, \delta_0, V_0) \equiv [n(M_1, \delta_1)V_0]^i \; [1 + \Xi_i(M_1, \delta_1, V_0)], \qquad (3.12)$$

where $q(\delta_1, \delta_0, V_0)$ is the probability that the overdensity is δ_0 when smoothed on scale V_0 and that it is less dense than δ_1 for all $V > V_0$. Notice that the first factorial moment of the halo distribution (Ξ_1) corresponds to the mean bias relation from Mo & White (1996).

Therefore, the mean of the bias relation from the MW model and the phenomenological modification of the Sheth & Lemson (1999) formula for the variance[1], are given by

$$\langle N \rangle = \int dm \; N(m, \delta_1 \mid M, \delta_0) \qquad (3.13)$$

and

$$\sigma^2 = \langle N(N-1) \rangle + \langle N \rangle - \langle N \rangle^2, \qquad (3.14)$$

where:

$$\langle N(N-1) \rangle = \int dm_1 \, dm_2 \; N(m_1, \delta_1 \mid M, \delta_0) \; N(m_2, \delta_1 \mid M - m_1, \delta')(1 + A\bar{\xi}_2), \quad (3.15)$$

$N(m, \delta_1 \mid M, \delta_0)$ denotes the average number of haloes of mass m identified at a given epoch z_1 [with a critical overdensity for collapse $\delta_1 = \delta_c(1+z_1)$] in an uncollapsed spherical region of comoving volume V with mass M and overdensity δ_0, and δ' is the mass density contrast of the fraction of the volume not occupied by the m haloes. The additional term $(1 + A\bar{\xi}_2)$ in the expression for the variance accounts for the contribution from mass clustering and has been constructed as the simplest function of the variance of the mass distribution with the property of having high values in overdense regions and of being unity in homogeneous regions. As it will be shown below, a good fit to the

---

[1]Only the spherical model is used here because a consistent implementation of the ellipsoidal model into the phenomenological model for the variance is not straightforward.

simulation data can be achieved by choosing $\overline{\xi}_2$ to be the second order moment of the mass distribution on the scale in consideration. In this case, we can write the term $A\overline{\xi}_2 = A\overline{\xi}_m(z_1) \approx AD^2(z_1)\overline{\xi}_m(0)$, where $D(z)$ is the linear growth factor normalized to one at $z = 0$. The constant A is to be calibrated by simulations.

3.3 Test by N-Body Simulations

3.3.1 Numerical Data

For this study the spatial distribution of dark matter particles as well as of dark haloes from the ΛCDM version of the high resolution GIF N-body simulations have been used (for details see Kauffmann et al. 1999). These simulations have 256^3 particles in a grid of 512^3 cells, with a gravitational softening length of $20\,h^{-1}$kpc. In the ΛCDM case, the simulation assumes $\Omega = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$. The initial power spectrum has a shape parameter $\Gamma = 0.21$ and is normalized so that the *rms* of the linear mass density in a sphere of radius $8\,h^{-1}$Mpc is $\sigma_8 = 0.9$. The simulation box has a side length $L = 141\,h^{-1}$Mpc, and the mass of each particle is $M_p = 1.4 \times 10^{10}\,h^{-1}M_\odot$.

The halo catalogues have been created by the GIF project (Kauffmann et al. 1999) using a friends-of-friends group-finder algorithm to locate virialized clumps of dark matter particles in the simulations outputs. They used a linking length of 0.2 times the mean interparticle separation and the minimum allowed mass of a halo is 10 particles. In what follows, the mass of a halo is represented by the number of particles it contains.

Galaxy catalogues constructed from the same simulations are also used. The catalogues are limited to model galaxies with masses greater than $\sim 2 \times 10^{10}\,h^{-1}M_\odot$. For further details about these catalogues and the galaxy formation models used in their construction see Kauffmann et al. (1999).

In order to study the halo-mass bias relation in detail, the conditional probabilities $P_V(N \mid \delta_m)$ have been estimated for several halo samples. The data used in the analysis correspond to the dark matter particle positions and the dark matter halo position catalogues at various redshifts $z = 3.5,\ 3.0,\ 2.7,\ 2.1,\ 1.5,\ 1.0,\ 0.5$ and 0 as well as the catalogues of the positions of the present time ($z = 0$) descendants of haloes already present in the above mentioned catalogues. Hereafter the redshift of identification of the haloes shall be denoted as z_1 whereas the redshift of analysis (i.e. the epoch at which the

counts-in-cells analysis is performed) as z_0. For example, a case with $z_0 = 0$ and $z_1 = 3$ means that haloes are identified at redshift 3 while the counts-in-cells are estimated for their central particles at redshift 0. From each of these halo and particles catalogues several mass ranges of the haloes have been studied. The halo-mass ranges analyzed go from samples containing all the haloes in the catalog (i.e $M_h > 10$ particles), to samples containing only the most massive haloes in the catalog, through a wide variety of subsets containing either small haloes, very massive haloes or intermediate mass haloes.

The catalogues of the present day descendants of haloes identified at earlier epochs have been created as follows: at a given redshift $z > 0$ the central particles of the haloes present in the corresponding catalog are identified and their positions traced forward to the present time. In this way, I create a catalog of the present time positions of the central particles of haloes identified at an earlier epoch, where the properties of the original haloes are transfered to the particle, in order to use them as sampling parameters.

The algorithm proposed by Szapudi et al. (1999), which allows an accurate determination of the probability function in a relatively short time, is applied to estimate the counts-in-cells on a grid of 256^3 cells. With this grid one obtains the counts-in-cells distribution at the scales $\ell = 1/256,\ 1/128,\ \cdots,\ 1/2$ times the side length of the simulation box.

The conditional probability to find N haloes in a cell of volume V given that the local mean mass overdensity has a value between δ_m and $\delta_m + \Delta\delta_m$ is computed from the counts-in-cells through

$$P_V(N|\delta_m) = \frac{P(N, \delta_m)\,\Delta\delta_m}{P(\delta_m)\Delta\delta_m}, \tag{3.16}$$

where $P_V(N, \delta_m)$ is the joint probability for finding N haloes and a mass overdensity between δ_m and $\delta_m + \Delta\delta_m$ in a cell of volume V. $P_V(\delta_m)$ is the distribution function for the underlying mass density field.

3.3.2   The Form of the Conditional Probability

The exact form of the conditional probability function has been investigated for many redshift pairs (z_1, z_0) and ranges of the halo masses in the samples, as mentioned above. Without loss of generality and for convenience, only a few of the analyzed samples will be shown explicitly here. In addition, due to numerical limitations (e.g. there are too few haloes at very small scales) the results obtained for very small scales are too noisy to

be used reliably in this investigation and therefore are not included. Thus, the analysis will be restricted to the the counts-in-cells performed on volumes of cubical cells with side lengths 1/32, 1/16, 1/8, and 1/4 times the side length of the simulation box, which correspond in the GIF ΛCDM simulations to $\ell = 4.4,\ 8.8,\ 17.6$ and $35.2\,h^{-1}$Mpc in comoving units.

Conditional Probability from the Simulations

Figures (3.2)–(3.3) show some contour plots of the conditional probability functions obtained from the simulations. The samples shown correspond to haloes identified and analyzed at the present epoch as well as to present epoch descendants of haloes identified at redshift 1. For clarity, only samples of haloes with masses larger than 10 particles are shown.

From the figures it can be seen that, indeed, there is a scatter in the *halo-mass bias* relation and that this scatter does not seem to be proportional to the mean bias relation at a given mass density contrast (δ_m).

Mainly due to numerical limitations the conditional probability obtained from the simulations has some noise. From observing figure (3.2) one can see that the conditional probability is more noisy in the lowest contours than in the highest. That means that, for a given interval in the mass density contrast ($\delta_m,\ \delta_m + \Delta\delta_m$), the conditional probability increases the noise at the extremes, which is a consequence of the fact that the events in the tails of the distribution are rare and therefore can not be sampled efficiently with the present numerical limitations. On the other hand, from the same figure it can be noticed that the conditional probability presents a large amount of noise in regions of high mass density contrast. Therefore any interpretation of the conditional probability in regions of very high mass density should be taken carefully.

The noise at the tails of the conditional probability, at a given δ_m, decreases as the redshift of identification of the haloes increases. Although the same happens to the noise at high mass overdensity regions, there is still a large amount of it in these regions.
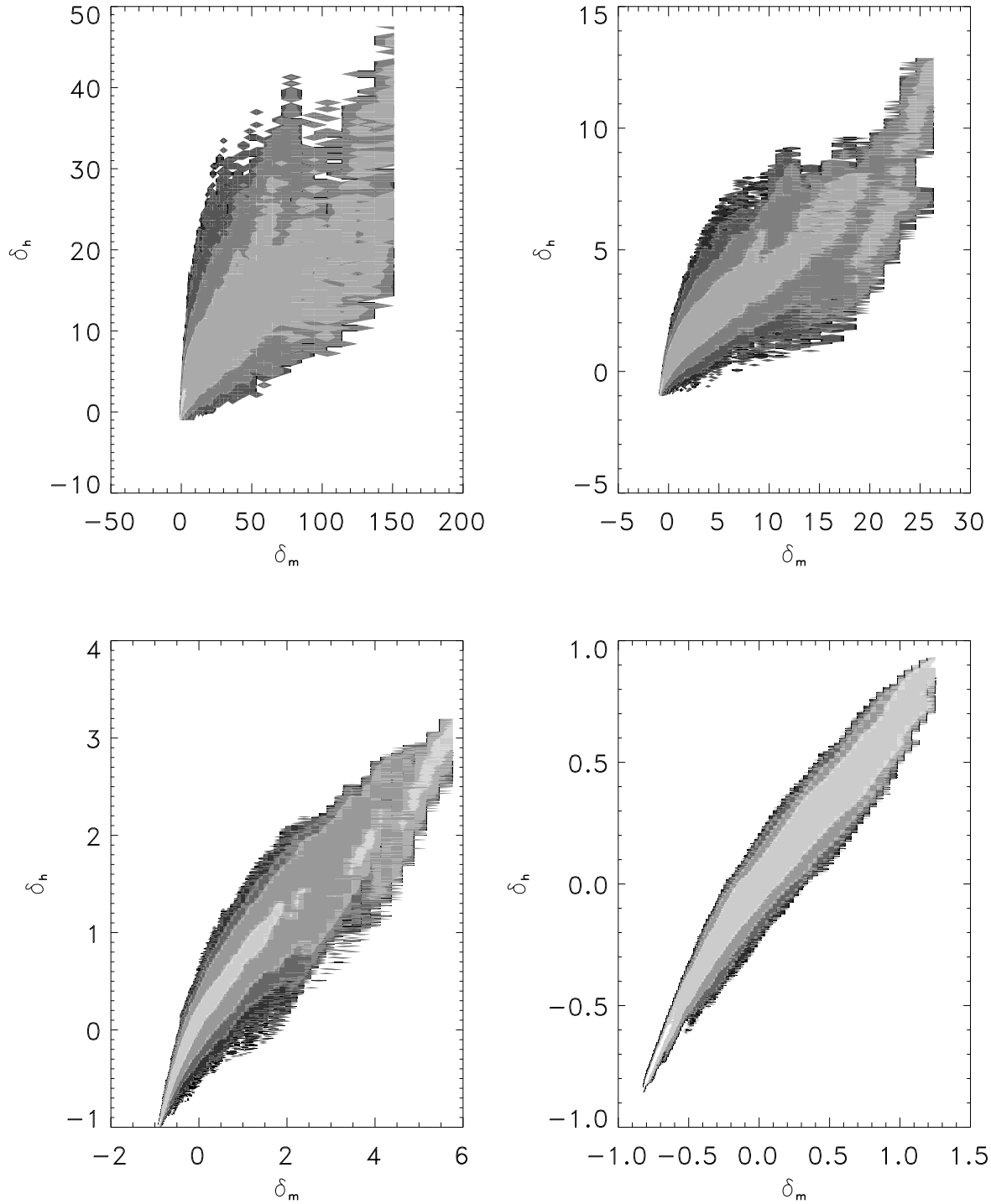
Figure 3.2:   Contour plot of the conditional probability obtained from the numerical data (i.e. mass and halo catalogues) at the present epoch. The halo sample contains all the haloes with masses larger than 10 particles. The sampling volumes correspond to cubical cells of side $\ell = 4.4$, 8.8, 17.6, 35.2 h^{-1}Mpc, from left to right and top to bottom. The contours are plotted for logarithmic levels of the conditional probability, ranging from 10^{-6} to 1 times the maximum value of the probability function.
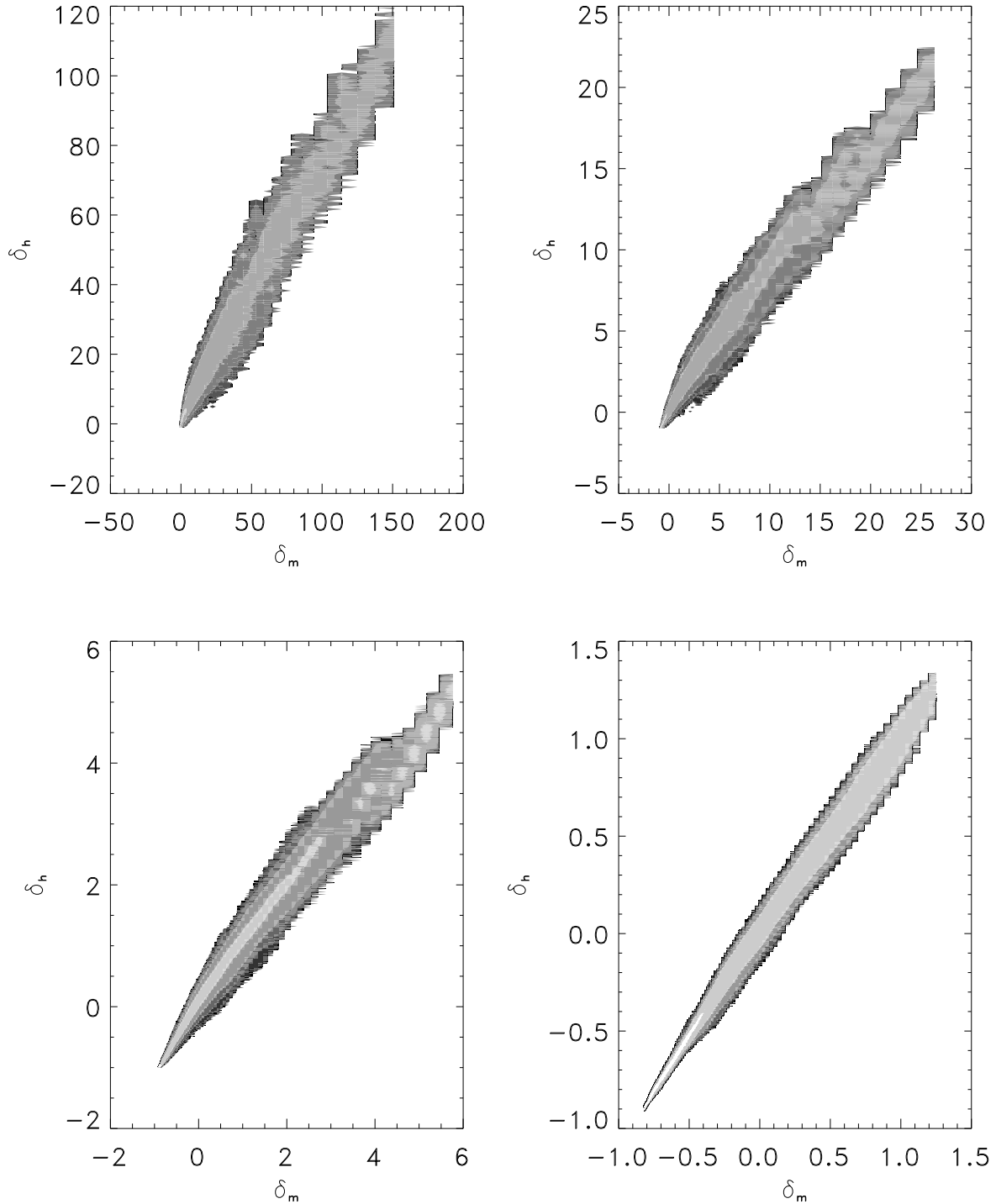
Figure 3.3: Contour plots of the conditional probability obtained from the numerical data. The halo samples contain all the haloes with masses larger than 10 particles. The sampling volumes correspond to cubical cells of side $\ell = 4.4$, 8.8, 17.6, 35.2 h^{-1}Mpc, from left to right and top to bottom. The contours are plotted for logarithmic levels of the conditional probability, ranging from 10^{-6} to 1 times the maximum value of the probability function. The plots correspond to haloes identified at redshift $z_1 = 1$ and analyzed at the present epoch $z_0 = 0$.

The conditional probability functions for the present time descendants of haloes iden-tified at earlier epochs [figure (3.3)] are, in general, less noisy than in the case of haloes identified and analyzed at the same epoch. Nevertheless, the regions of high mass over-density still having a large amount of noise. One particular feature of the conditional probability of the present day descendants of early times haloes is that its scatter is lower in comparison with the corresponding quantities for haloes analyzed at the identification epoch.

Looking for the Best Fitting Function

Now let us investigate the analytical form of the bias relation (conditional probability function) for several samples of haloes. Once the conditional probabilities for several halo samples have been estimated from the simulations and knowing some general features of them, a numerical fit of the analytical distribution functions introduced in section 3.2.1 to the numerical probability functions has been performed. For that the counts-in-cells of the corresponding spatial distribution of dark matter particles as well as of the distribution of dark matter haloes in the corresponding samples have been estimated. From these counts-in-cells it is easy to get the halo-mass joint probability function and the mass distribution function, both binned in δ_m, needed to obtain the conditional probability function, as given by equation (3.16).

The problem of finding the best fit to the numerical data boils down to the minimiza-tion of a smooth nonlinear sum of squares

$$\text{Minimize} \quad F(\mathbf{x}) \equiv \frac{1}{2} \sum_{i=1}^{m} [y_i - f_i(\mathbf{x})]^2, \tag{3.17}$$

where m is the number of data points, $\mathbf{x}$ is a vector of the fitting parameters, y_i is the i-*th* value of the numerical function and $f_i(\mathbf{x})$ is the corresponding value of the analytical function to fit. $F(\mathbf{x})$ is usually known as the χ^2 of the fit. The measure of the goodness of a given fit is thus quantified by means of the corresponding χ^2, with lower values of it meaning a better fit of the function to the data.

As already stated in section 3.2.1, the different parameters, characterizing each of the analytical functions, are taken as free fitting parameters (i.e. there is not a predefined relationship between the parameters and the mass overdensity δ_m nor with the mean of

the halo distribution function). In the following the results of the fitting process for a selected set of halo samples will be presented.

The first quantity to be analyzed corresponds to χ^2 itself. Figures (3.4) and (3.5) present χ^2 obtained from the fit of each of the four analytical functions under analysis to the numerical data, as a function of the mass density contrast δ_m. The samples shown correspond to present epoch haloes with masses larger than 10 particles and to present epoch descendants of haloes with masses larger than 10 particles identified at redshift 1.

In the case of present epoch haloes (figure (3.4)), it is evident that the Poisson distribution (dotted lines) is by far not a good description of the numerical conditional probability for the sample and sampling scales shown, while the Gaussian distribution appears as a strong candidate to be the best fitting function. χ^2 values from the fitting of the Lognormal and Thermodynamic distributions to the data are less regular than χ^2 corresponding to the Gaussian. Furthermore, only at the smallest scale (upper left plot) the Thermodynamic distribution function fits partially better than the Gaussian. Therefore one can conclude that the Gaussian distribution function describes well the numerical conditional probability from the samples shown in figure (3.4), and that the Poisson distribution is rather a poor descriptor of the same numerical conditional probability.

For samples of haloes at higher redshifts (not shown) it is observed that the behavior of the χ^2 from the fits is more or less the same as the one obtained for present epoch haloes. The only appreciable difference is that the Poisson distribution fits better to the numerical distribution as the redshift increases. However, it is still a poorer description for the numerical conditional probability in comparison to the Gaussian distribution.

Similarly, the results obtained from samples of the present epoch descendants of haloes already formed at high redshift (figures (3.5) and (3.5)) do not change appreciably with respect to the corresponding results in the case of haloes identified and analyzed at the same redshift.

Therefore, from the analysis of the behavior of the χ^2 obtained from the fitting of each of the four analytical functions to the conditional probability obtained from the simulations, it can be concluded that the Poisson distribution is rather a poor description for the conditional probability and that the Gaussian distribution is a better descriptor of the conditional probability. This is true for all the halo samples investigated.
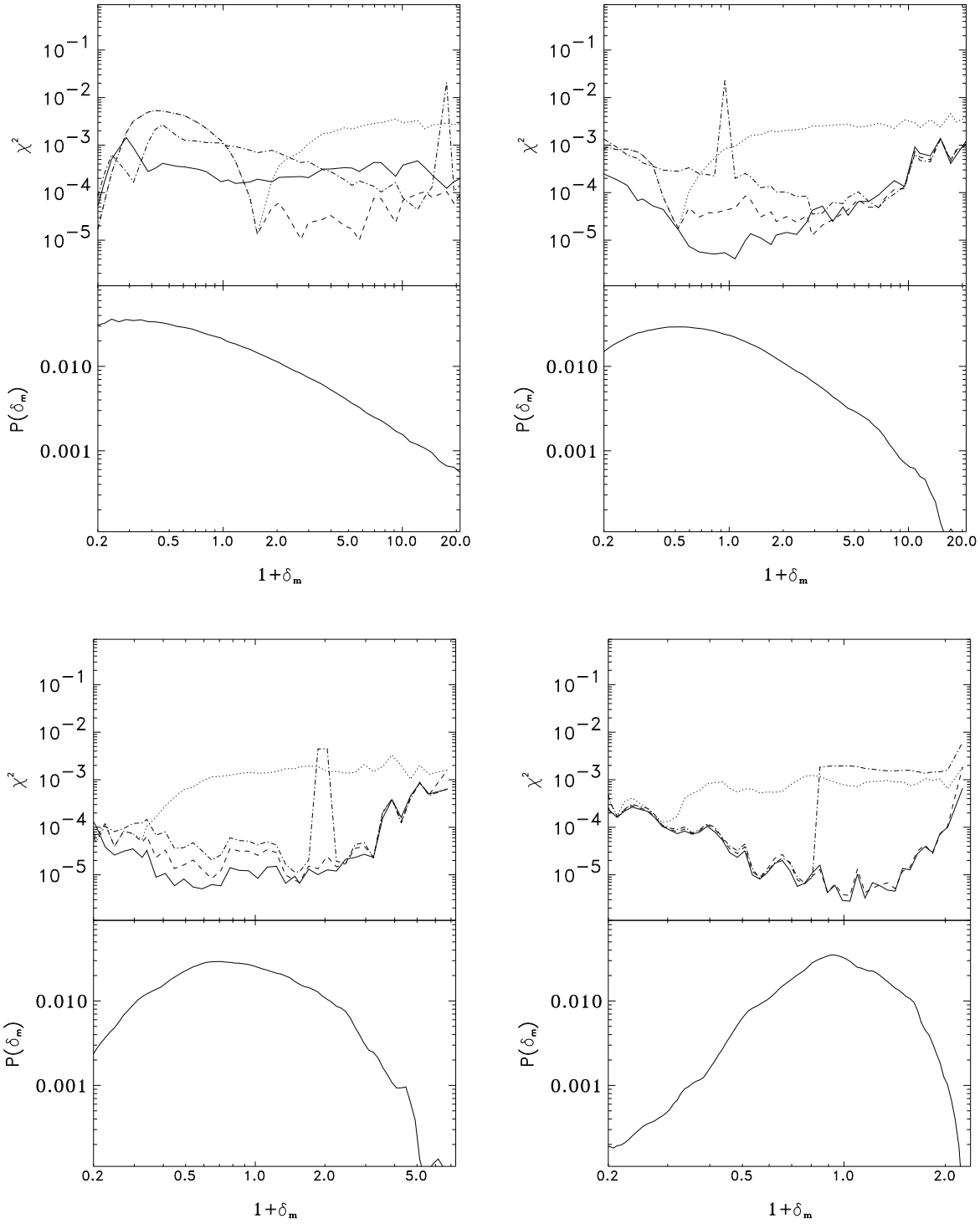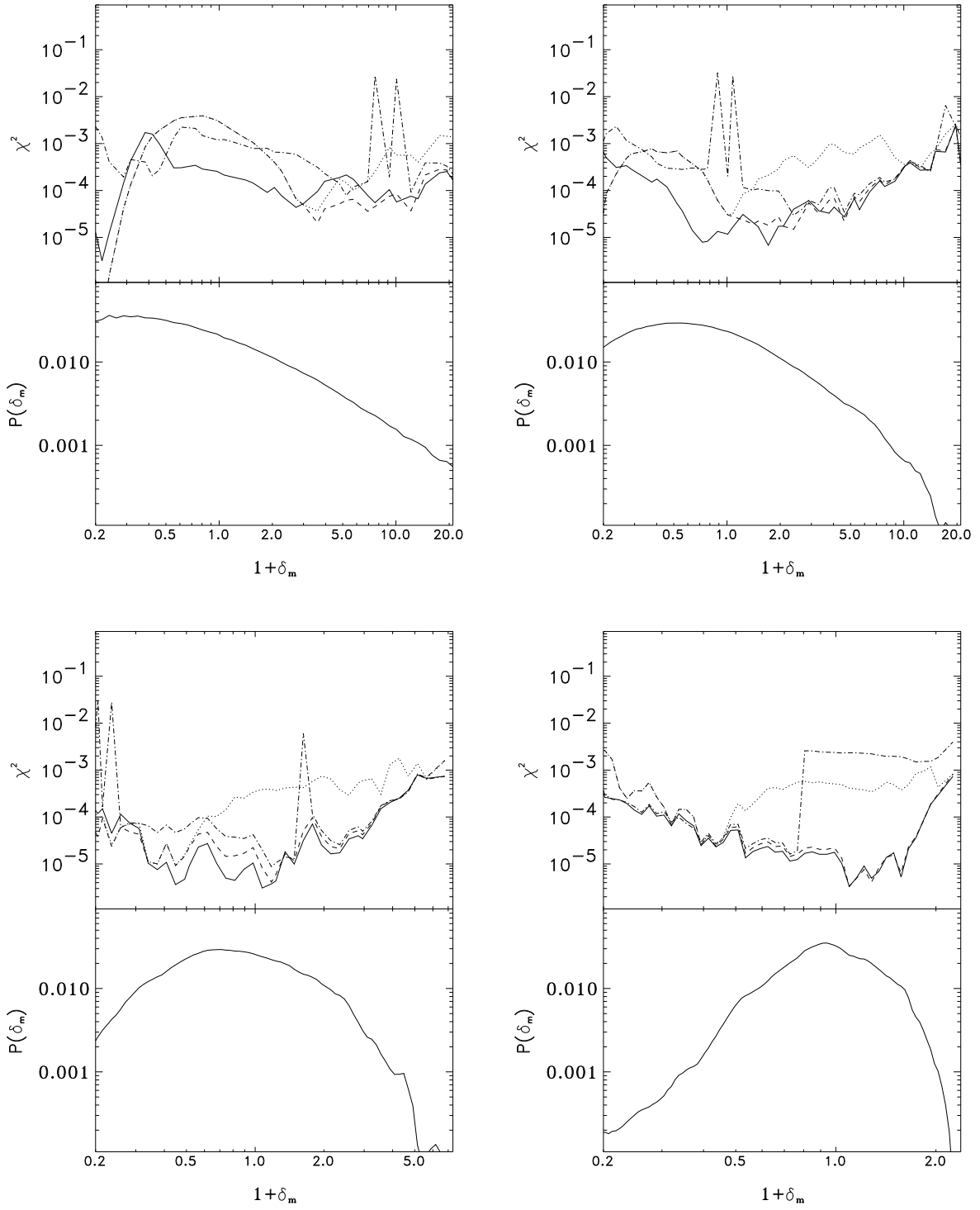
Figure 3.4: χ^2, as a function of the mass density contrast δ_m, from the fit to the conditional probability from the simulations of the Poisson (dotted-line), the Thermodynamic (dashed-line), the Lognormal (dash-dot-line) and the Gaussian (solid-line) distributions. The sample corresponds to all the present epoch ($z_1 = z_0 = 0$) haloes in the catalog ($M_h > 10$ particles). The corresponding mass probability functions are also plotted. The panels correspond to the scales $\ell = 4.4, \ 8.8, \ 17.6 \ $ and $35.2 \ h^{-1}$Mpc, from top to bottom and left to right.

Figure 3.5: χ^2, as a function of the mass density contrast δ_m, from the fit to the conditional probability obtained from the simulations of the Poisson (dotted-line), the Thermodynamic (dashed-line), the Lognormal (dash-dot-line) and the Gaussian (solid-line) distributions. The sample corresponds to all the present epoch ($z_1 = 1$, $z_0 = 0$) haloes in the catalog ($M_h > 10$ particles). The corresponding mass probability functions are also plotted. The panels correspond to the scales $\ell = 4.4$, 8.8, 17.6 and 35.2 h^{-1}Mpc, from top to bottom and left to right.

In addition, figures (3.6)–(3.9) present the conditional probability functions $P_V(N|\delta_m)$ obtained from the simulations at several representative values of the mass density contrast δ_m, along with the corresponding best fits of the Gaussian, Lognormal, Poisson and Thermodynamic functions. The halo samples shown correspond to present epoch haloes as well as to present day descendants of haloes identified at redshift 1. For clarity, only two different halo-mass ranges are shown for each redshift.

From figures (3.6)–(3.7) it can be confirmed that the Poisson model is in general a poor description of the present time conditional probability measured from the simulations, and that the Gaussian model is overall a good assumption. The Lognormal and Thermodynamic functions are a sort of intermediate functions, i.e they are not as poor descriptors for the conditional probability as the Poisson function is, but, on the other hand, they do not describe the features of the conditional function as good as the Gaussian function does. This result is valid for all the halo mass ranges under analysis and all the scales tested.

Similarly, from figures (3.8) (3.9) it can be seen, again, that the Poisson distribution is not a good descriptor for the conditional probability function obtained from the simulations, and that the Gaussian function is the best descriptor, among the four functions under analysis.

The results already shown, altogether, lead to the conclusion that, the Poisson distribution function does not describe well the conditional probability function of haloes, and that the Gaussian distribution is a better descriptor for the numerical function. The Lognormal and Thermodynamic functions are in between the Poisson and the Gaussian behavior. This conclusion is true for all the redshifts of identification and analysis investigated here, as well as for all ranges of the mass of the haloes in the samples and all the scales covered in this investigation.

3.3.3   The Mean and Variance of Halo-Mass Bias

In the last section it has been shown that the Gaussian distribution is a reasonable fit to the conditional probability function obtained from the numerical simulations. Therefore, let us now concentrate on the mean and the variance of the conditional probability (bias relation), which are the two quantities needed to completely specify a Gaussian distribution. In order to investigate the deviations of the bias relation from the Poisson
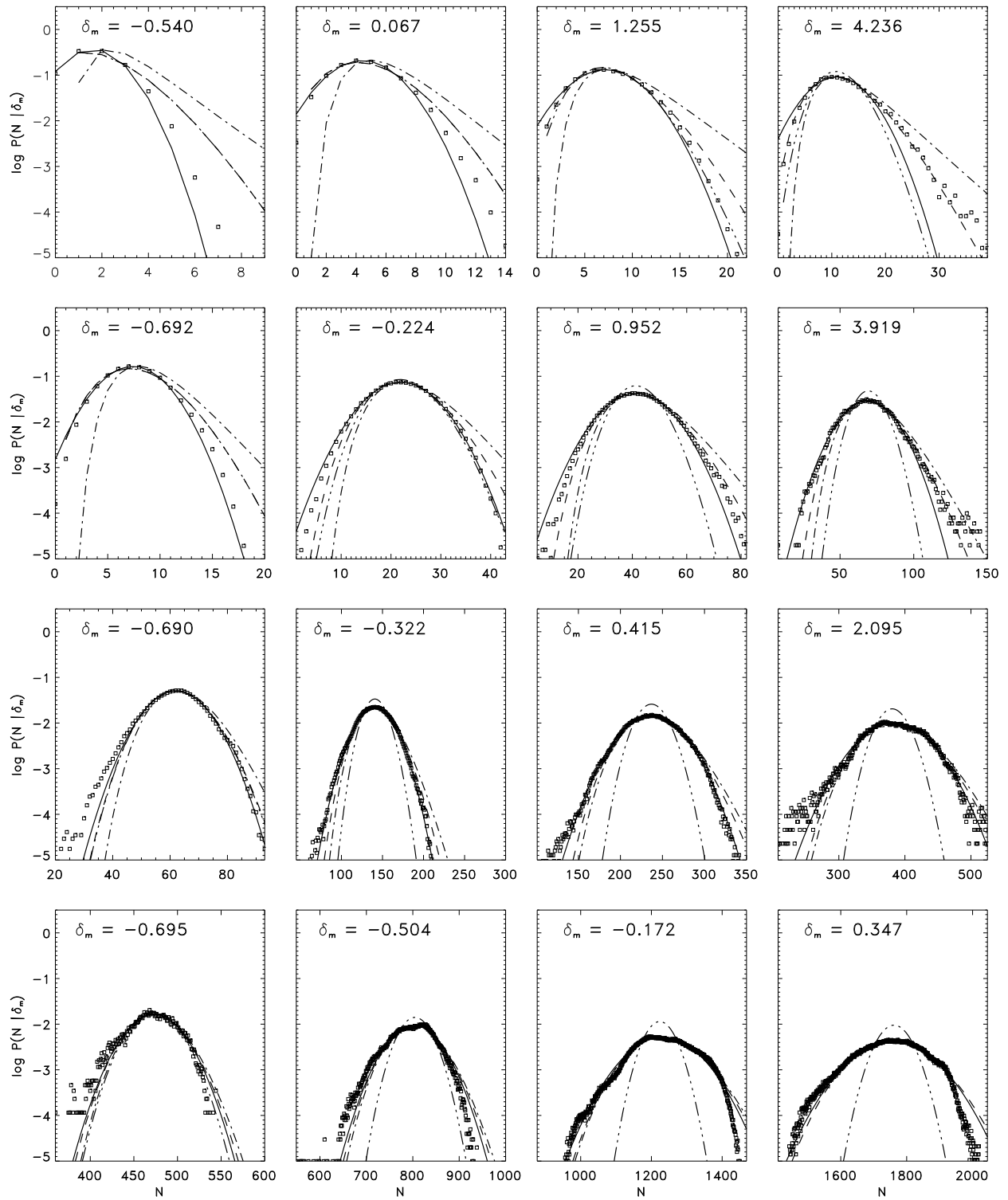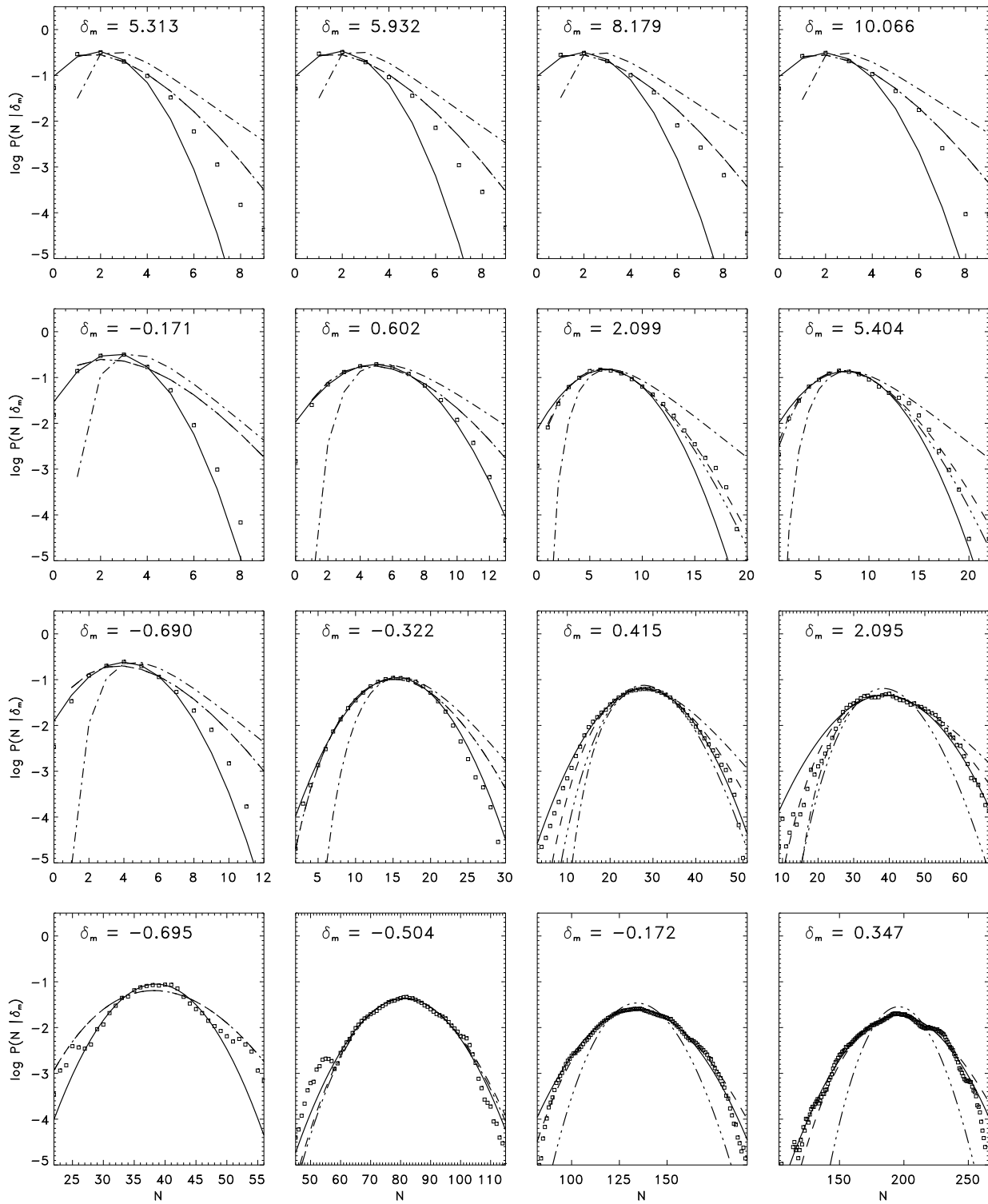
Figure 3.6: Comparison between the conditional probability measured from the simulations (squares) for present epoch haloes with masses greater than 10 particles and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales $\ell = 4.4,\ 8.8,\ 17.6,\ 35.2\ Mpc/h$, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.

Figure 3.7: Comparison between the conditional probability measured from the simulations (squares) for present epoch haloes with masses greater than 100 particles and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales $\ell = 4.4,\ 8.8,\ 17.6,\ 35.2\ Mpc/h$, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.
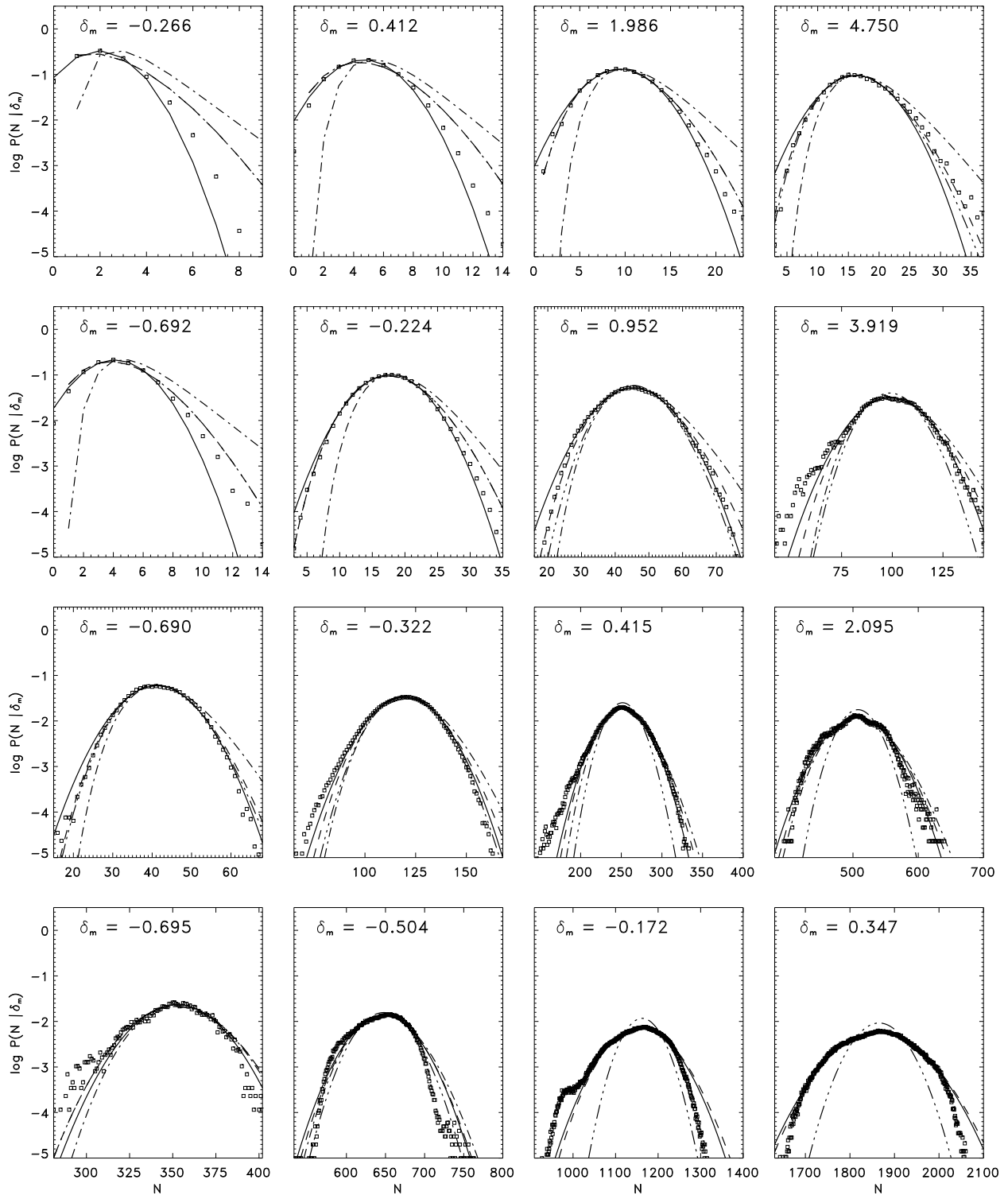
Figure 3.8: Comparison between the conditional probability measured from the simulations (squares) for present day descendants of haloes at redshift 1 with masses greater than 10 particles and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales 4.4, 8.8, 17.6, 35.2 Mpc/h, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.
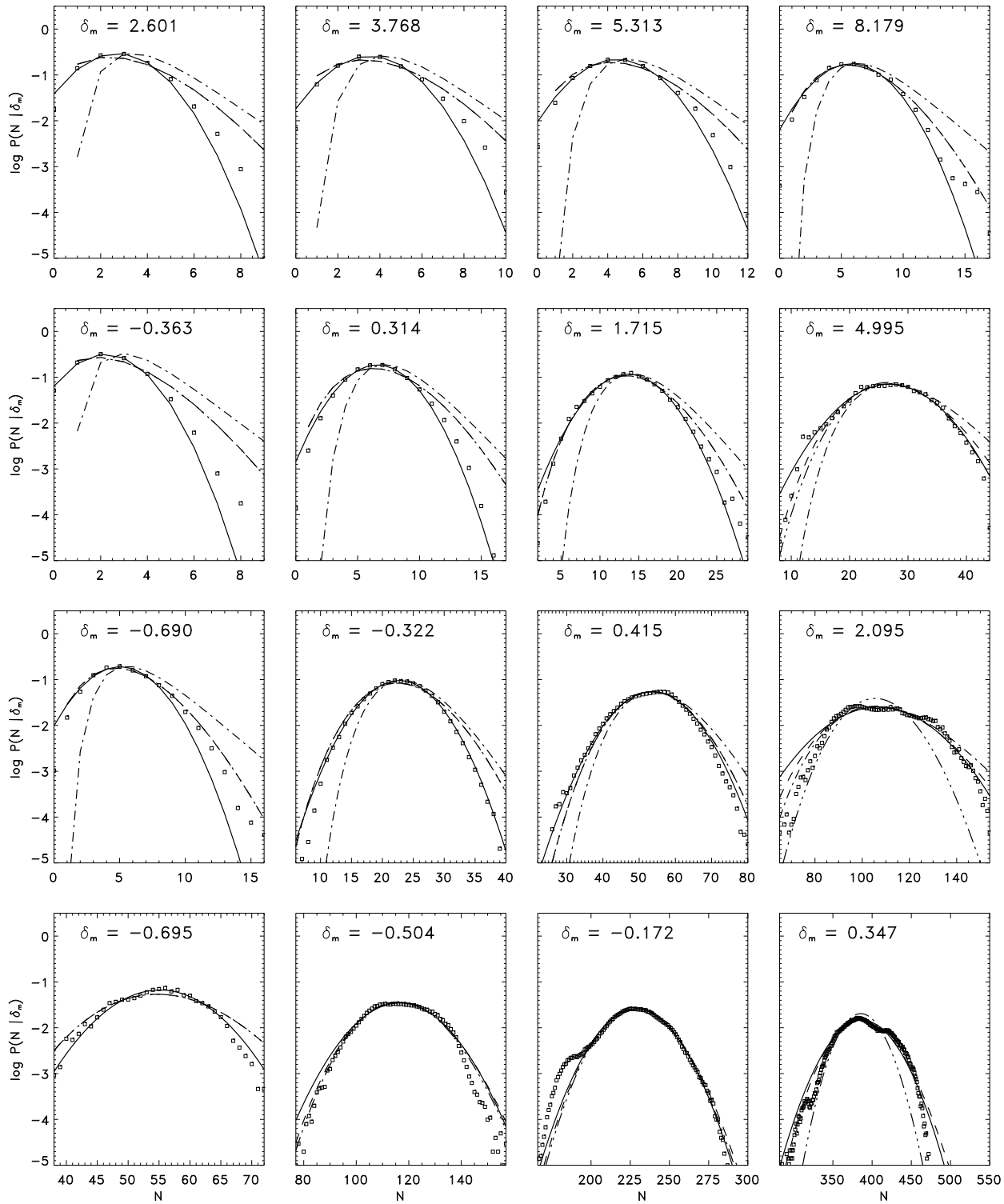
Figure 3.9: Comparison between the conditional probability measured from the simulations (squares) for present day descendants of haloes at redshift 1 with masses greater than 50 particles and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales 4.4, 8.8, 17.6, 35.2 Mpc/h, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.

distribution the ratio between the variance and the mean shall be used, rather than the variance itself. Let us consider the mean of the bias relation

$$1 + \delta_h \equiv \frac{\langle N | \delta_m \rangle}{\overline{n} V} \tag{3.18}$$

and the ratio between the variance and the mean

$$\frac{\text{variance}}{\text{mean}} \equiv \frac{\sigma^2}{\langle N | \delta_m \rangle}, \tag{3.19}$$

where $\delta_h \equiv \frac{N}{\overline{n} V} - 1$ is the number density contrast of haloes and $\overline{n}$ is their mean number density.

In the following the predictions from the theoretical models for the mean (Mo & White 1996) and for the variance (Sheth & Lemson 1999) of the halo-mass bias relation shall be compared with the corresponding numerical quantities obtained from the simulations. The theoretical models for the values of these quantities as functions of redshift, mass of the haloes, local mass density contrast and the particular cosmology, have been introduced in section 3.2.2.

The halo samples used in this analysis are mainly the same as in the last section. In order to investigate in detail the behavior of the ratio variance/mean, the results for halo samples in four representative mass ranges will be shown. They are: a) a sample of low mass haloes, b) a sample containing both low and high mass haloes, c) a sample of intermediate mass haloes and d) a sample of high mass haloes. The corresponding halo masses are shown in table 3.1.

Figures (3.10)–(3.14) show the results obtained from the simulations and from the model, for haloes at redshift ($z_1 = z_0 = 0$, 1, 3) as well as for the present epoch descendants of haloes already identified at these epochs. Observing the symbols in the plots, which correspond to the quantities obtained from the simulations, it can be noticed that the ratio of the variance to the mean of the bias relation shows a Poisson-like behavior (i.e. ~ 1) for low values of δ_m. This ratio becomes sub-Poisson (i.e. < 1) at intermediate values of δ_m, and super-Poisson (> 1) for high values of δ_m. The exact change of the variance/mean ratio with δ_m depends on halo mass: the sub-Poisson variance extends to higher values of δ_m for samples with higher halo masses. The volume-exclusion effect is reduced for the descendants of haloes identified at an earlier epoch and the variance/mean

	$z_1 = 0$	$z_1 = 1$	$z_1 = 3$
a)	$M_h = 20\text{--}30$ part.	$M_h = 20\text{--}30$ part.	$M_h = 20\text{--}30$ part.
	$M_h/M_\star = 0.03\text{--}0.04$	$M_h/M_\star = 0.62\text{--}0.92$	$M_h/M_\star = 100\text{--}150$
b)	$M_h = 20\text{--}2000$ part.	$M_h = 20\text{--}2000$ part.	$M_h = 20\text{--}600$ part.
	$M_h/M_\star = 0.03\text{--}2.85$	$M_h/M_\star = 0.62\text{--}61.5$	$M_h/M_\star = 100\text{--}300$
c)	$M_h = 200\text{--}800$ part.	$M_h = 200\text{--}800$ part.	$M_h = 50\text{--}100$ part.
	$M_h/M_\star = 0.28\text{--}1.14$	$M_h/M_\star = 6.15\text{--}24.6$	$M_h/M_\star = 250\text{--}500$
d)	$M_h > 800$ part.	$M_h > 800$ part.	$M_h > 200$ part.
	$M_h/M_\star > 1.14$	$M_h/M_\star > 24.6$	$M_h/M_\star > 1000$

Table 3.1: Ranges of halo masses corresponding to the samples shown in figures (3.10)–(3.14). a) sample of low mass haloes, b) sample containing both low and high mass haloes, c) sample of intermediate mass haloes and d) sample of high mass haloes. $M_\star$ is defined by $\sigma(M_\star) = 1.68$.

ratio approaches the Poisson value for the descendants of haloes selected at early times (see figures (3.13)–(3.14)).

The curves in figures (3.10)–(3.14) show the predictions from the models for the mean and the variance of the bias relation. The mean bias relations given by the simulations are well described by the model of Mo & White (1996), confirming earlier results.

The value of the constant A in equation (3.14) is calibrated using the mean and variance of the bias relation for the present epoch haloes in the simulations. The best approximation found for this constant is $A = 0.05$ and has been obtained as given by the fit of the model predicted ratio variance/mean for present-day haloes to the corresponding quantity from the simulations.

The behavior of the variance/mean ratio is also reasonably well reproduced by the model. Thus, sub-Poisson variance can be caused by halo exclusion while the super-Poisson variance at high δ_m may be explained by the clustering of mass at the time of halo identification. The model for the variance begins to fail at very high values of δ_m. But since cells with such high densities are only a tiny fraction of all cells, this failure might not be very important.

In the case of samples of present epoch haloes (figure 3.10)), the model for the variance of the bias relation has been found to work remarkably good. However, as it can be seen from the figure, the sample containing both small and massive haloes fails slightly at the smallest scale shown ($\ell = 4.4\ h^{-1}\mathrm{Mpc}$). This feature might be due to a cross-correlation
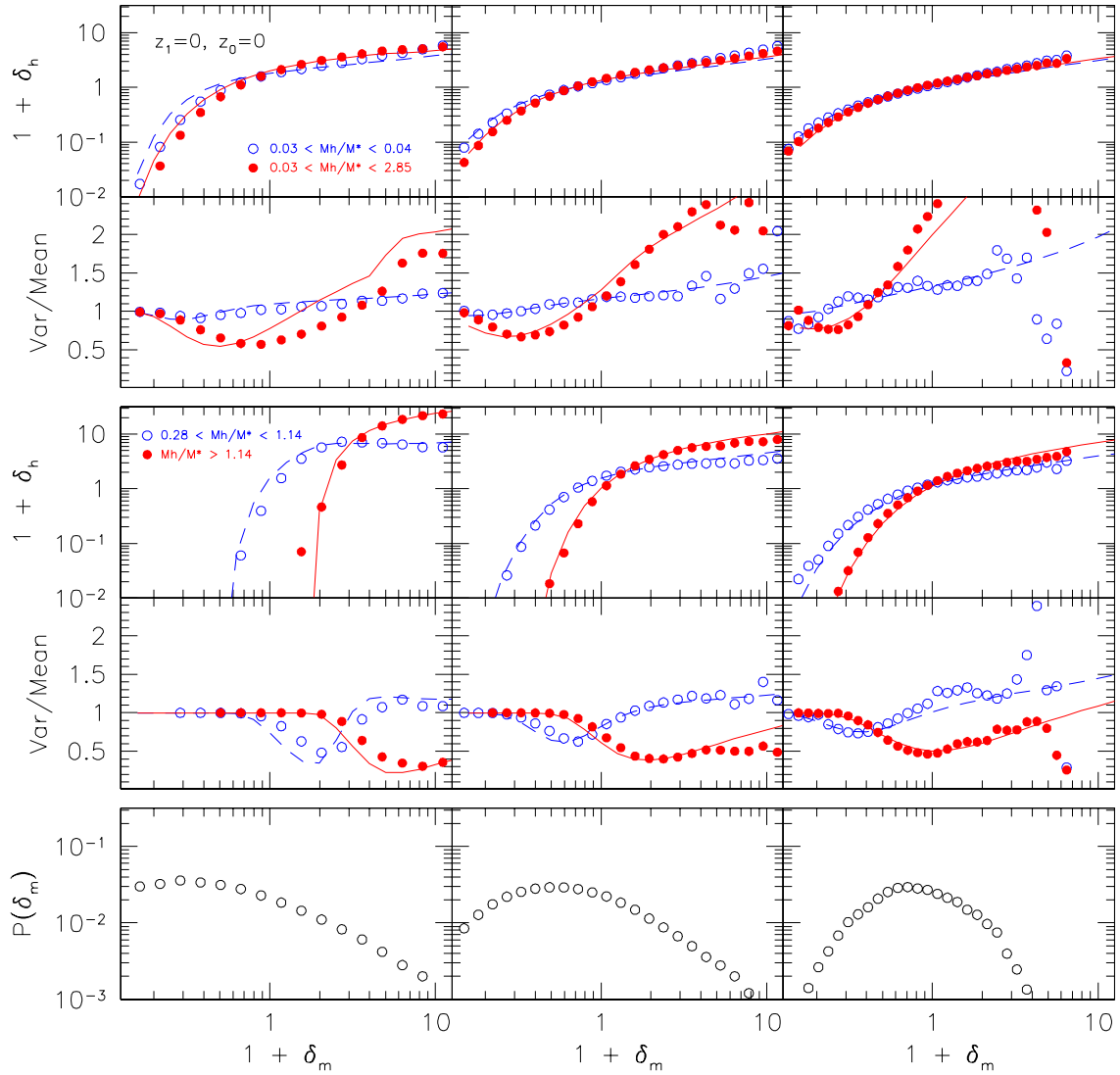
Figure 3.10: Theoretical predictions from the MW model for the mean and from the proposed phenomenological modification of the (Sheth & Lemson 1999) model for the variance of the bias relation (lines) compared with the corresponding quantities obtained from the GIF ΛCDM simulations (symbols). The columns correspond, from left to right, to the cell sizes $l = 4.4$, 8.8 and 17.6 Mpc/h. The two upper panels show the mean of the bias relation and the ratio between the variance and the mean of the bias relation for the ranges of halo masses indicated in the respective labels. The dashed and solid lines show the theoretical predictions corresponding to the numerical data represented by the open and filled circles, respectively. The mass probability function at the respective scales is shown in the lowest panel. The sample corresponds to haloes identified and analyzed at the present epoch.
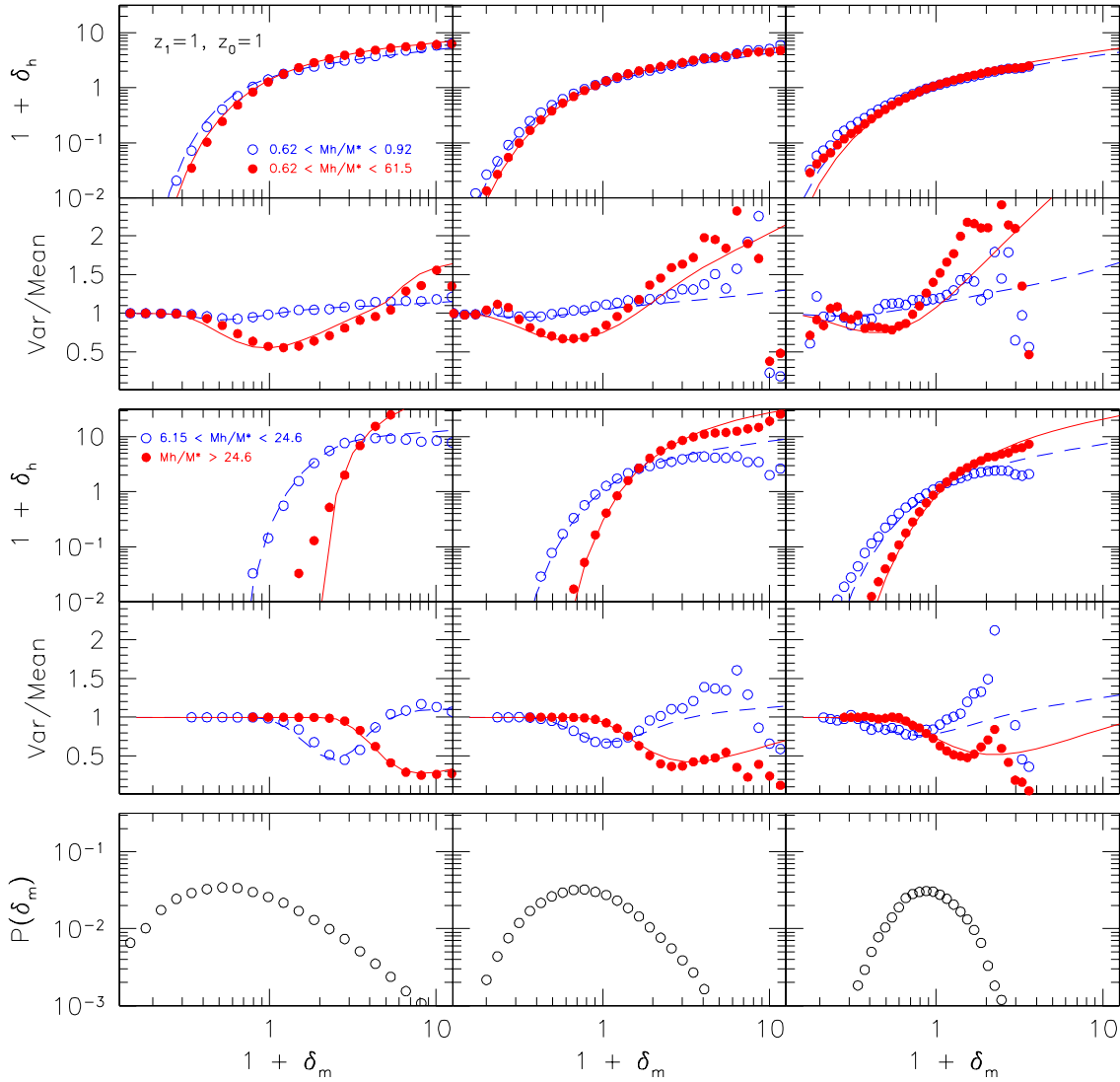
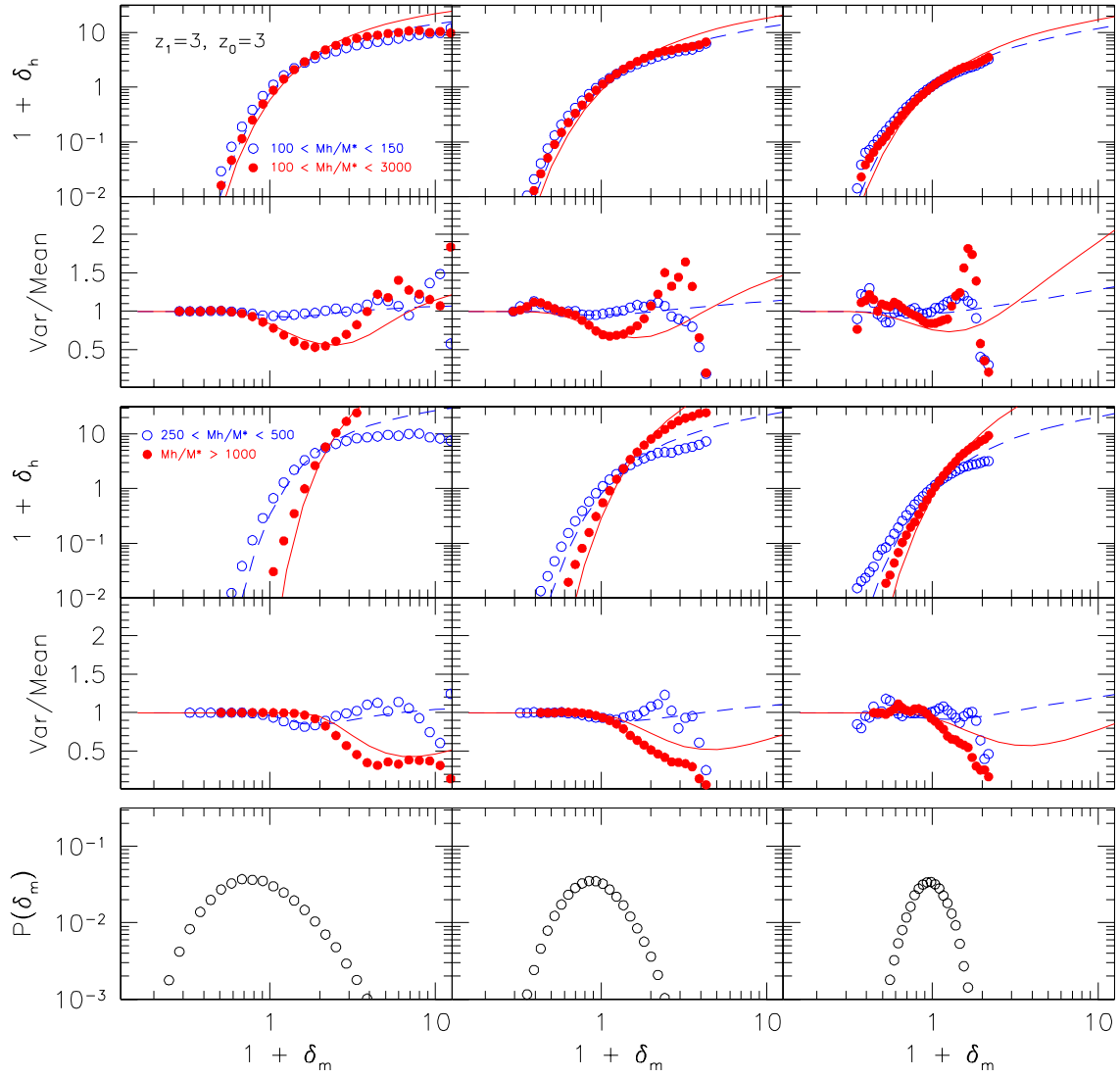Figure 3.11: Same results as shown in figure (3.10) but for haloes identified and analyzed at redshift $z = 1$.

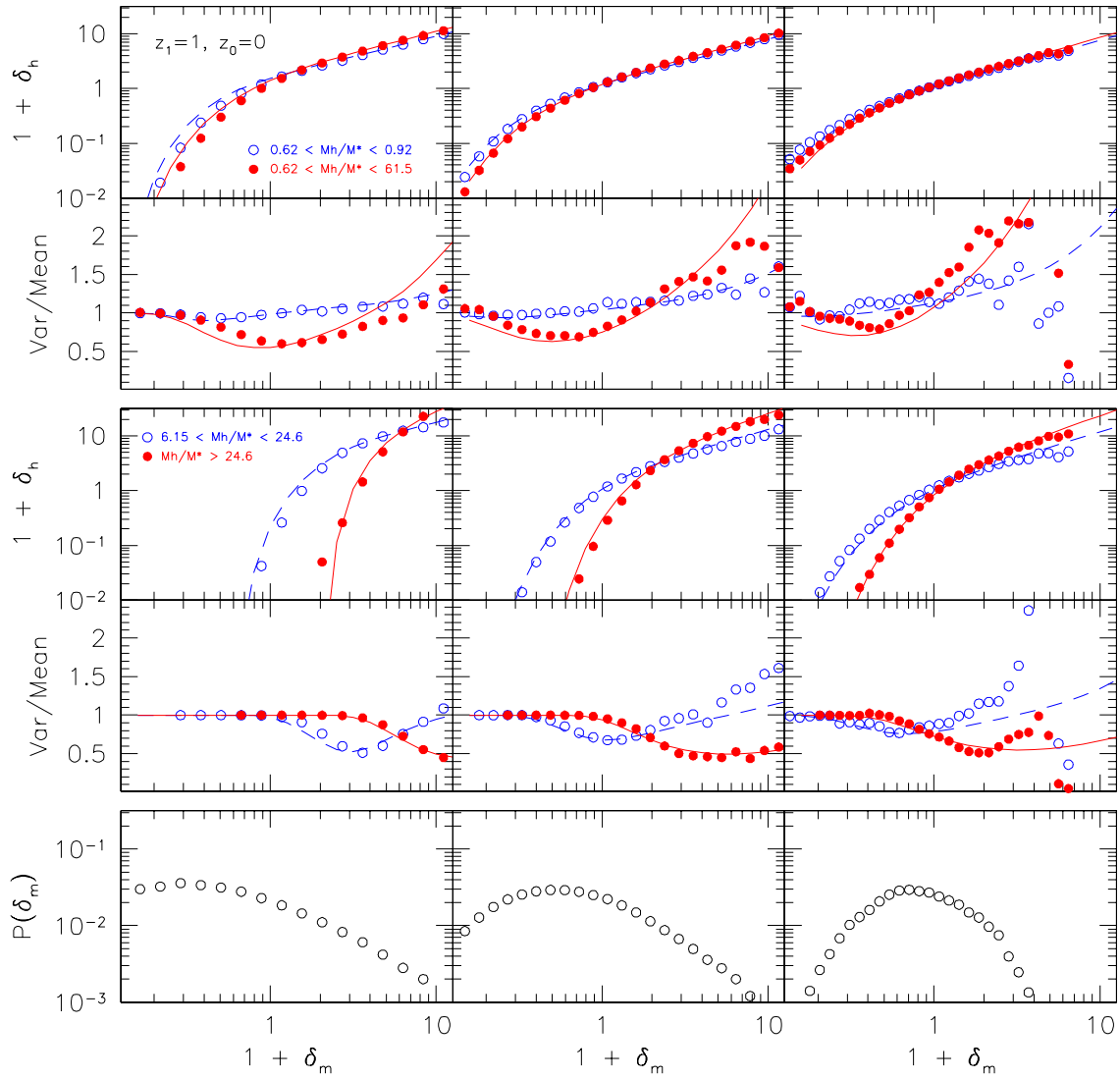Figure 3.12: Same results as shown in figure (3.10) but for haloes identified and analyzed at redshift $z = 3$.

Figure 3.13: The same as in figure (3.10) but for the present epoch descendants of haloes already formed at $z = 1$.
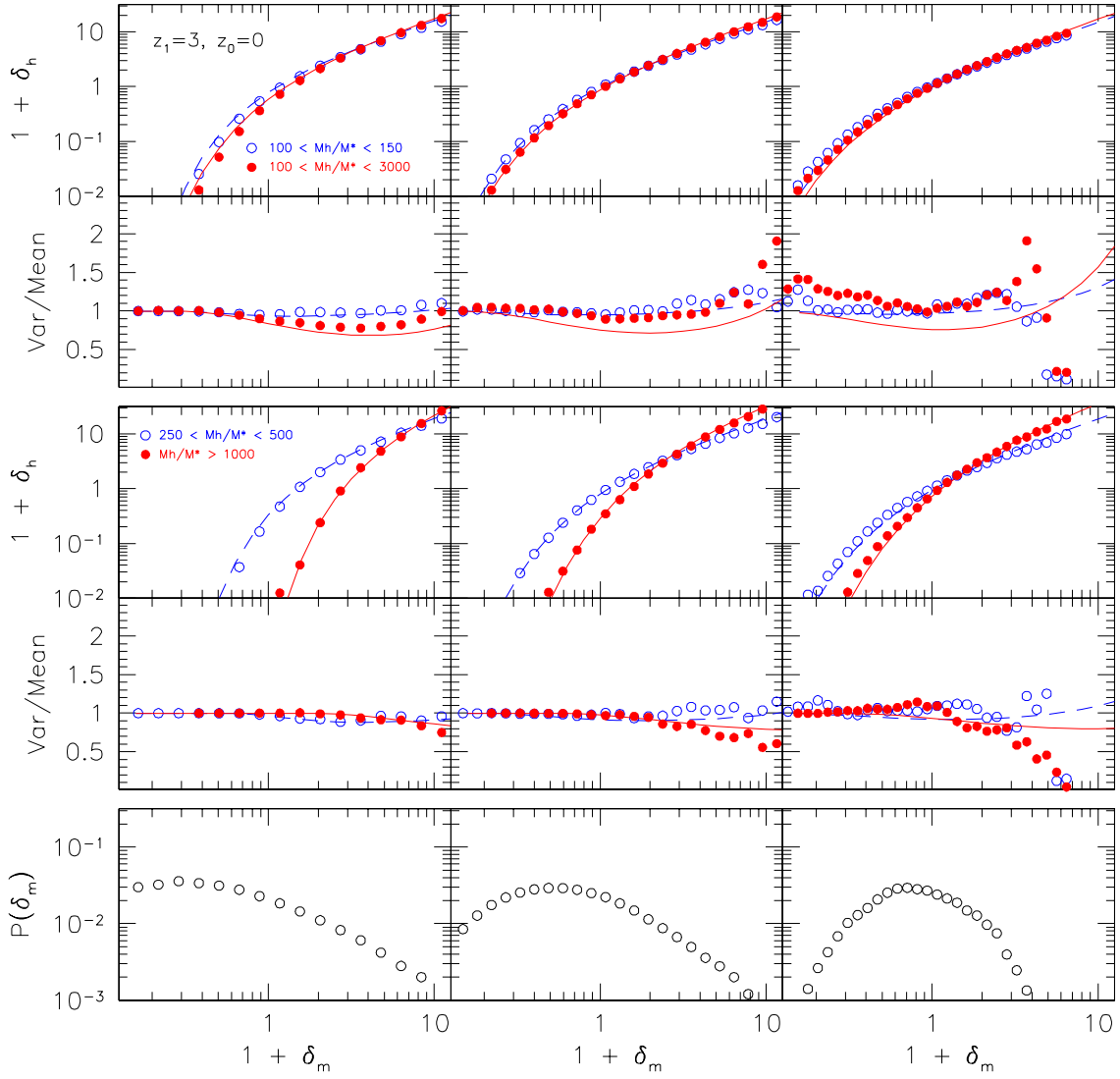
Figure 3.14: The same as in figure (3.10) but for the present epoch descendants of haloes already formed at $z = 3$.

term between low mass and high mass haloes. There is also a large disagreement between the model predictions and the numerical data at high overdensity regions, which is particularly noticeable at the largest scale shown in the plots ($\ell = 17.6\ h^{-1}\text{Mpc}$). Due to the large amount of noise in the conditional probability at these high-overdensity regions, it is not clear whether the model fails or simply the numerical data are no longer useful in these regions.

From figures (3.11)–(3.12) it can be observed that the model for the variance of the bias relation for haloes at high redshift ($z_1 = z_0 = 1$, and 3) presents an overall performance which is more or less similar to the one found for present epoch haloes. The main difference is that, in the case of high-redshift haloes, the model seems to work better at small scales and that it systematically under-predicts the ratio variance/mean at high-overdensity regions, specially at large scales.

In the case of the present time descendants of haloes identified at earlier epochs (figures (3.13)–(3.14)) the model works also remarkably well in all the scales and halo samples. At very high overdensity regions the model starts to fail, but, as it has been already discussed, it is not clear yet whether the model, the simulations data, or both are responsible for this failure. In the particular sample containing present day descendants of very low mass haloes at redshift 3 (figure (3.14)), the model seems to under-predict the value of the ratio variance/mean. This might be interpreted as due to an overestimation of the role of the volume exclusion effect, which is expected to decrease as the redshift of identification of the parent haloes increases.

Summarizing, it has been found that the model of Mo & White (1996) describes well the mean of the bias relation obtained from the simulations, which is a confirmation of earlier results. The proposed extension to the model of Sheth & Lemson (1999) has been found to describe remarkably well the variance of the bias relation obtained from numerical simulations.

3.3.4 The Count-in-Cell Function of Dark Haloes

An additional test that can be performed on the model of the halo-mass bias [i.e. a Gaussian conditional probability function with the mean and variance given by equations (2)-(4)] consists in reconstructing the counts-in-cells function for haloes using equation (3.1). In order to reconstruct the counts-in-cells of haloes (halo probability function),

the probability function of the mass ($P_V(\delta_m)$) obtained from the simulations has been used together with the conditional probability given by the theoretical model. Although theoretical models for the mass distribution function exist in the literature [e.g. the model of Sheth (1998) based on an excursion set approach, and the Lognormal model used in Coles & Jones (1991)], they are not used here, since the purpose of this analysis is only to test the model for the halo-mass.

Since the probability functions obtained from the simulations are quite noisy at very high values of δ_m and the model predictions in this regime may fail, the computations are truncated at a preset high value of δ_m [$\delta_m^{max} = 10$ at the scales $l = 4.4$ and $l = 8.8\ h^{-1}\ Mpc$, and $\delta_m^{max} = 3$ at $l = 17.6\ h^{-1}\ Mpc$], which corresponds to the low-probability tail of the mass probability function, as can be seen clearly in the lower panel in figure (3.10). For comparison the halo count-in-cell functions using a Poissonian form for the conditional function , with the mean given by equation the model of Mo & White (1996) (see equation (3.13)), is also reconstructed.

Figure (3.15) shows the comparison between the reconstructed halo count-in-cell functions for present-day haloes containing more than 10 particles with the corresponding functions obtained directly from the simulations. The plots shown correspond to the scales $\ell = 4.4,\ 8.8,\ and\ 17.6\ h^{-1}$Mpc. Since the reconstructed halo counts-in-cells have been obtained from a mass counts-in-cells truncated at very high values of the mass density contrast (δ_m), the halo counts-in-cells from the simulations are also re-estimated taking into account the mass truncation. From the plots it can be seen that the semi-analytically reconstructed probability function of the haloes matches remarkably well the corresponding distribution function obtained from the simulations, after truncation. The halo count-in-cell functions reconstructed using a Poissonian form for the conditional probability function depart from the corresponding numerical values in the low-probability, high density tail.

The halo count-in-cells functions reproduced through this approach might be used to estimate the high-order moments, such as skewness and kurtosis, of halo distributions. However, since a truncation in the mass distribution function has to be applied and the truncation value is introduced rather arbitrarily, this application is not promising before the model for the bias relation is improved at very high overdensity regions and numerical simulations with higher resolutions are available to test it.
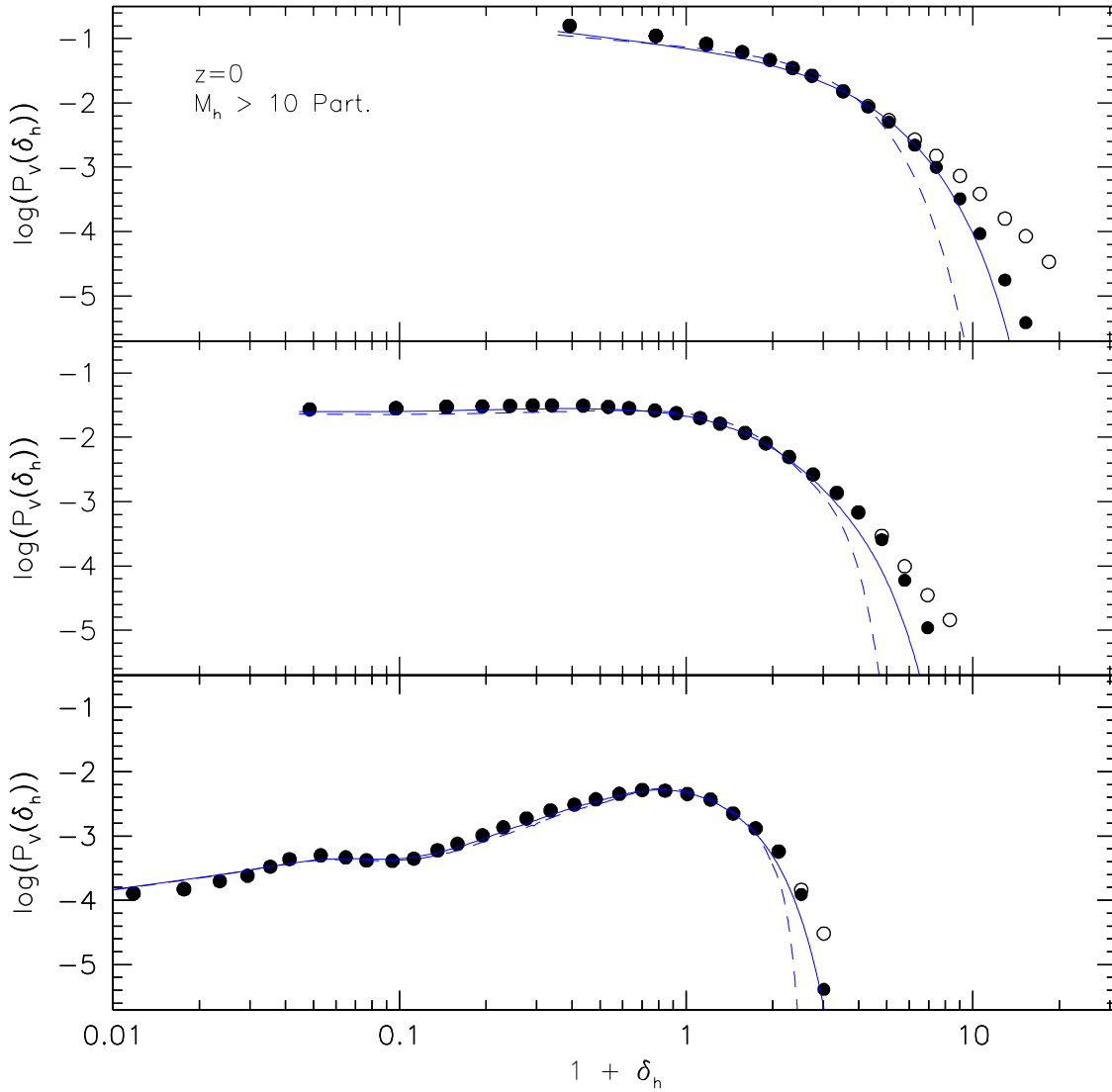
Figure 3.15: Halo count-in-cell functions for a sample of present day haloes with masses greater than 10 particles. The circles correspond to the probability function obtained from the simulations and the lines to the semi-analytically reconstructed count-in-cell function using spherical collapse approach. The solid and the dashed lines show the reconstructed functions using a Gaussian and a Poissonian form for the conditional probability function, respectively. The filled circles correspond to the simulated mass count-in-cell functions obtained from the mass and conditional probability functions truncated at high values of the mass density contrast. The boxes correspond, from top to bottom, to the scales $l = 4.4,\ 8.8,\ 17.6\ h^{-1}\ Mpc$.

3.3.5 Stochasticity in Galaxy Bias

So far only the relation between the mean and the variance of the halo-mass bias relation has been investigated. Nevertheless, it is very interesting to study the bias relation between galaxies and the underlying mass distribution. Thus, in order to investigate the stochastic nature of the galaxy bias and the possible deviations from Poisson of the variance of the galaxy-mass bias relation, the conditional probabilities for galaxies have been estimated using several galaxy catalogues created from the GIF simulations [for details see Kauffmann et al. (1999)] and the mass distributions from the same simulations. The catalogues contain model galaxies with masses greater than $\sim 2 \times 10^{10} h^{-1} M_{\odot}$.

The process applied to obtain the conditional probability from the numerical data is the same as in the case of dark matter haloes.

Figures (3.16)–(3.18) show the conditional probabilities from the model galaxy catalogues at redshift 0, 1 and 3. It can be seen that the galaxy conditional probability is very similar to the halo conditional probability.

The fitting process done for the halo conditional probability is performed also on the galaxy conditional probability with similar results; that is, that the galaxy conditional probability is better described by a Gaussian function than by a Poissonian one. The comparison between the numerical conditional probability of model galaxies and the best fits of the analytical functions, for several representative values of the mass overdensity δ_m is shown in figures (3.19)–(3.21)

The mean and variance of the bias relation between model galaxies and the underlying mass has been estimated for samples of model galaxies at the present epoch as well as at redshift 1. Figure (3.22) shows the results obtained at the cubical sampling volumes characterized by the scales $\ell = 4.4,\ 8.8,\ and\ 17.6\ h^{-1}$Mpc. Interestingly, the variance/mean ratio in the galaxy-mass bias relation also exhibits significant sub-Poissonian behavior, implying that the effect of volume exclusion is also important for the spatial distribution of galaxies. One possible reason for this is that many of the galaxy-sized haloes may host only one galaxy and the galaxy distribution inherits a considerable fraction of the exclusion effects from the distribution of their host haloes. If this result is also true in the case of real galaxies, that is that the (real) galaxy-bias relation is not described by a Poisson distribution, then there are important implications of this fact for the interpretations of galaxy clustering, as will be discussed at the end of the chapter.
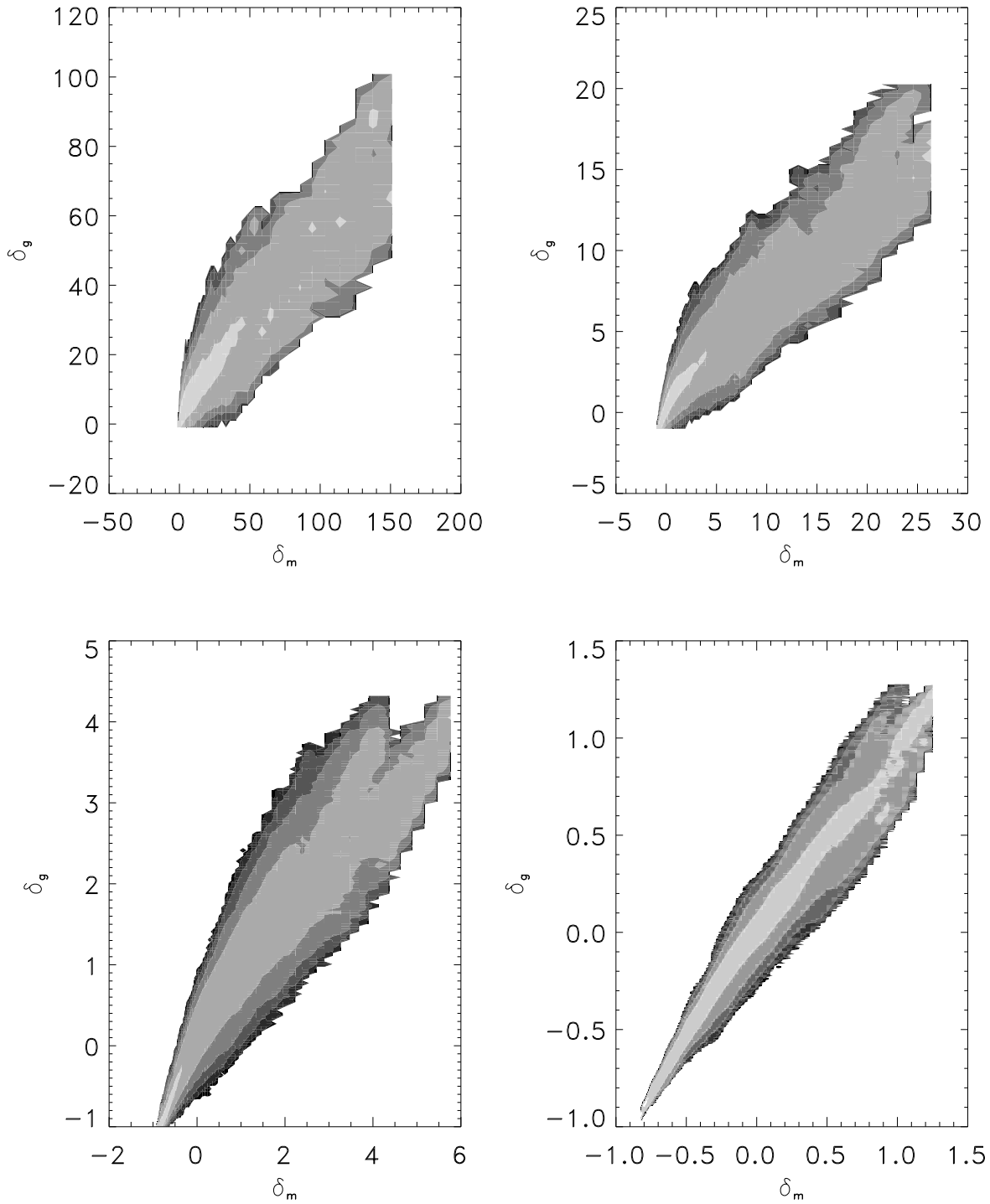
Figure 3.16:   Contour plot of the conditional probability obtained from the numerical data (i.e. mass and galaxy catalogues) at the present epoch. The sampling volumes correspond to cubic cells of side $\ell = 4.4$, 8.8, 17.6, 35.2 h^{-1}Mpc, from left to right and top to bottom. The contours are plotted for logarithmic levels of the conditional probability, ranging from 10^{-6} to 1 times the maximum value of the probability function.
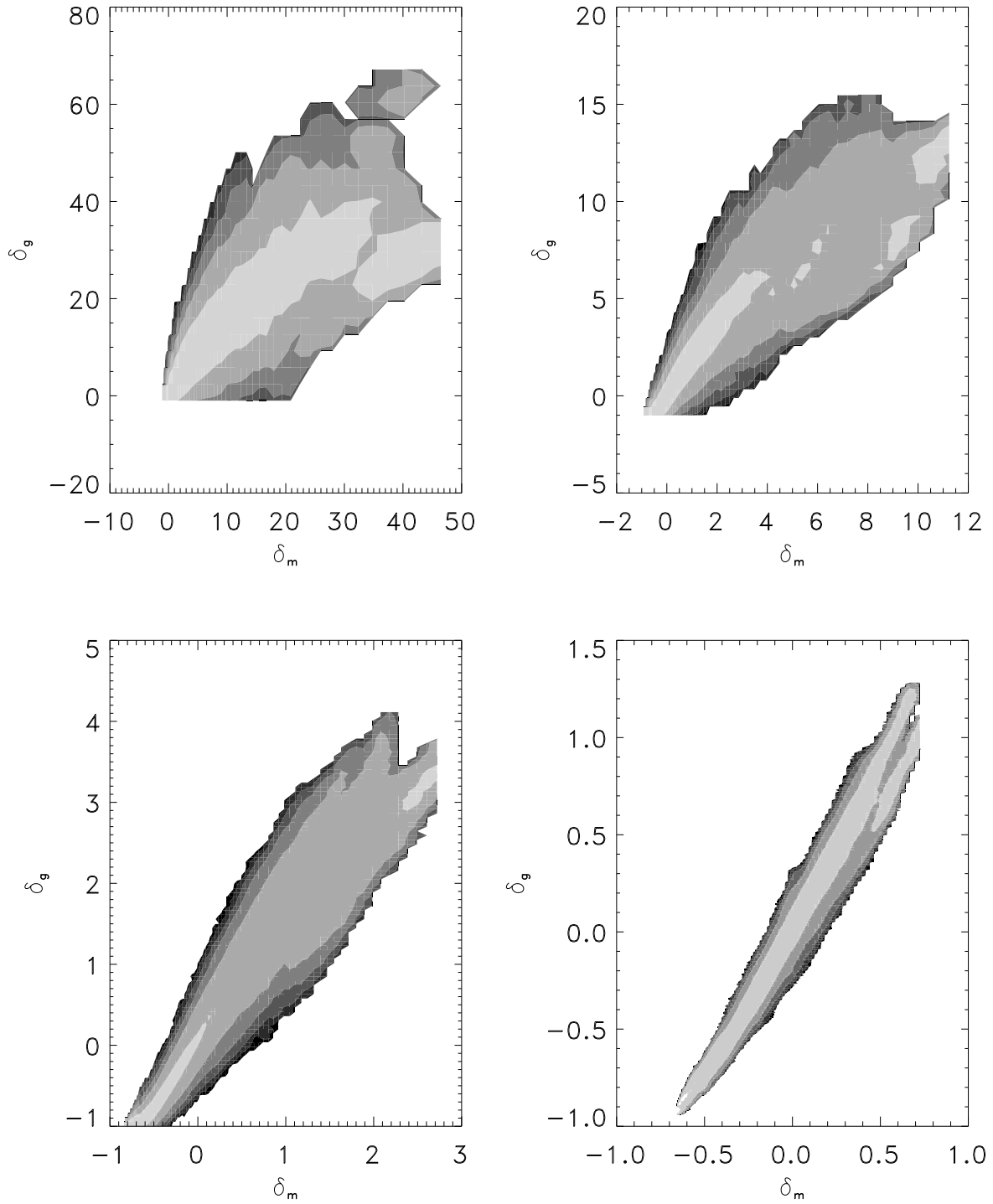
Figure 3.17: Contour plot of the conditional probability obtained from the numerical data (i.e. mass and galaxy catalogues) at redshift 1. The sampling volumes correspond to cubic cells of side $\ell = 4.4,\ 8.8,\ 17.6,\ 35.2\ h^{-1}$Mpc, from left to right and top to bottom. The contours are plotted for logarithmic levels of the conditional probability, ranging from 10^{-6} to 1 times the maximum value of the probability function.
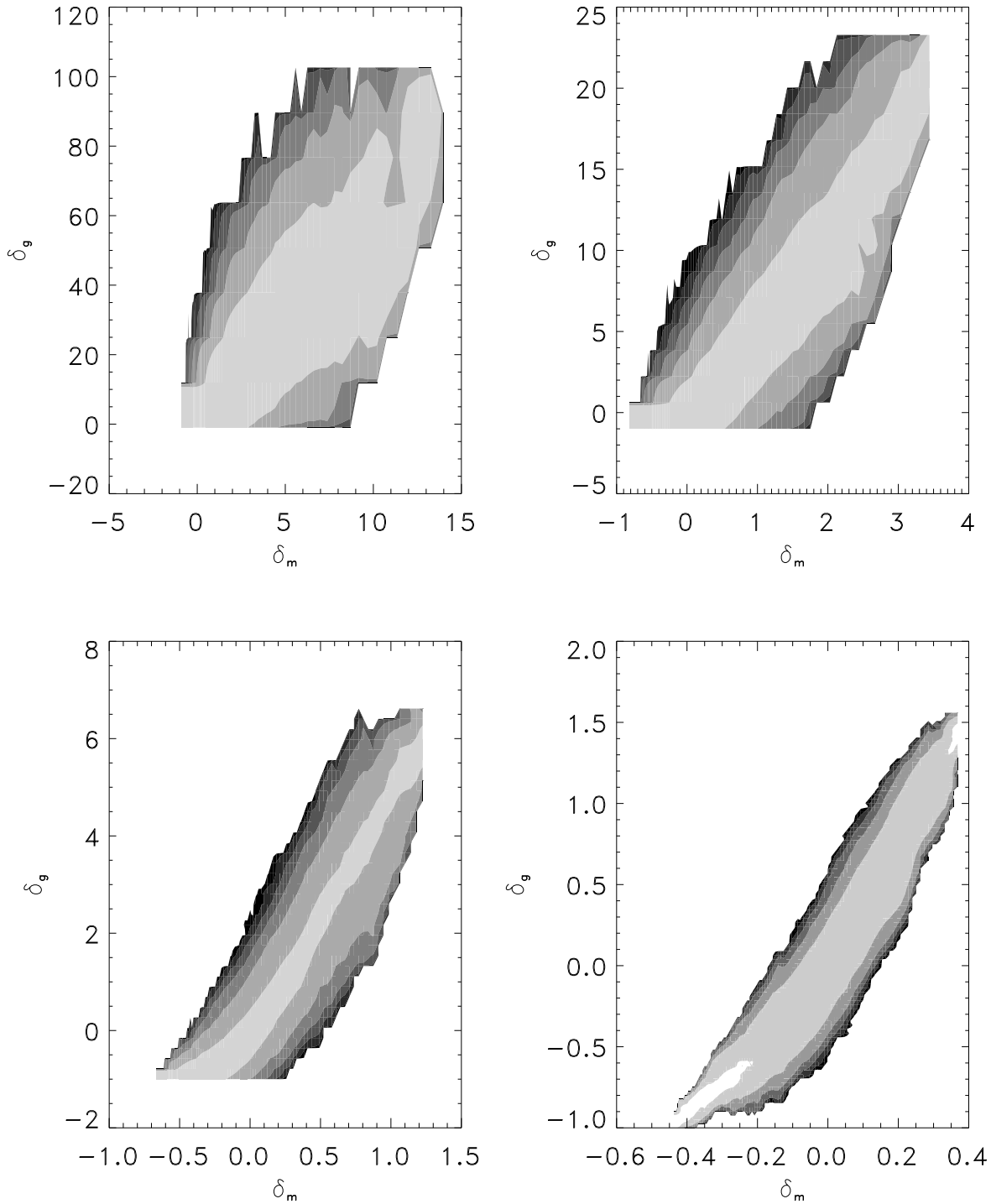
Figure 3.18: Contour plot of the conditional probability obtained from the numerical data (i.e. mass and galaxy catalogues) at redshift 3. The sampling volumes correspond to cubic cells of side $\ell = 4.4$, 8.8, 17.6, 35.2 h^{-1}Mpc, from left to right and top to bottom. The contours are plotted for logarithmic levels of the conditional probability, ranging from 10^{-6} to 1 times the maximum value of the probability function.
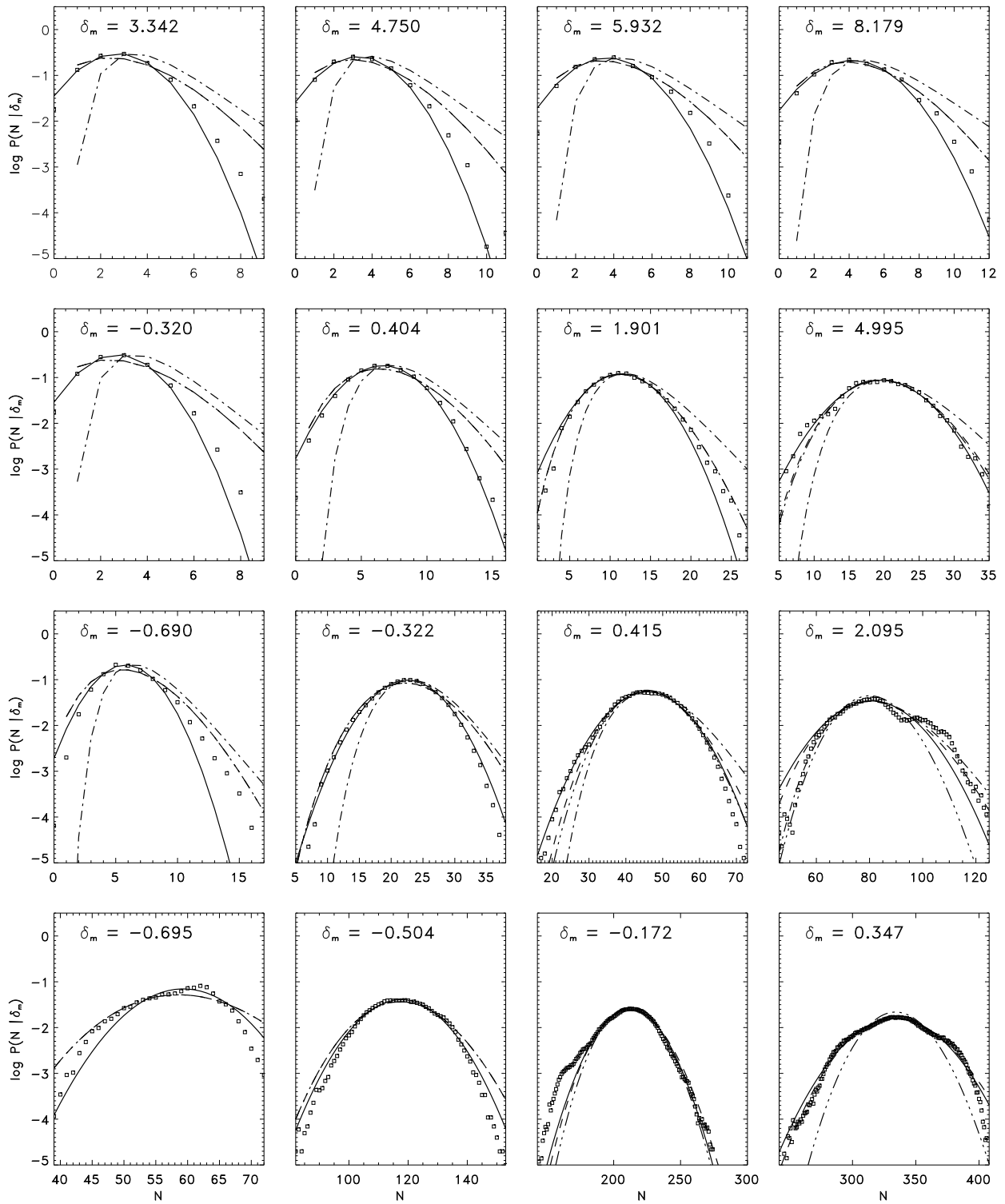
Figure 3.19: Comparison between the conditional probability measured from the numerical data (squares) for present day model galaxies and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales 4.4, 8.8, 17.6, 35.2 Mpc/h, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.

Figure 3.20: Comparison between the conditional probability measured from the numerical data (squares) for model galaxies at redshift 1 and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales 4.4, 8.8, 17.6, 35.2 Mpc/h, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.
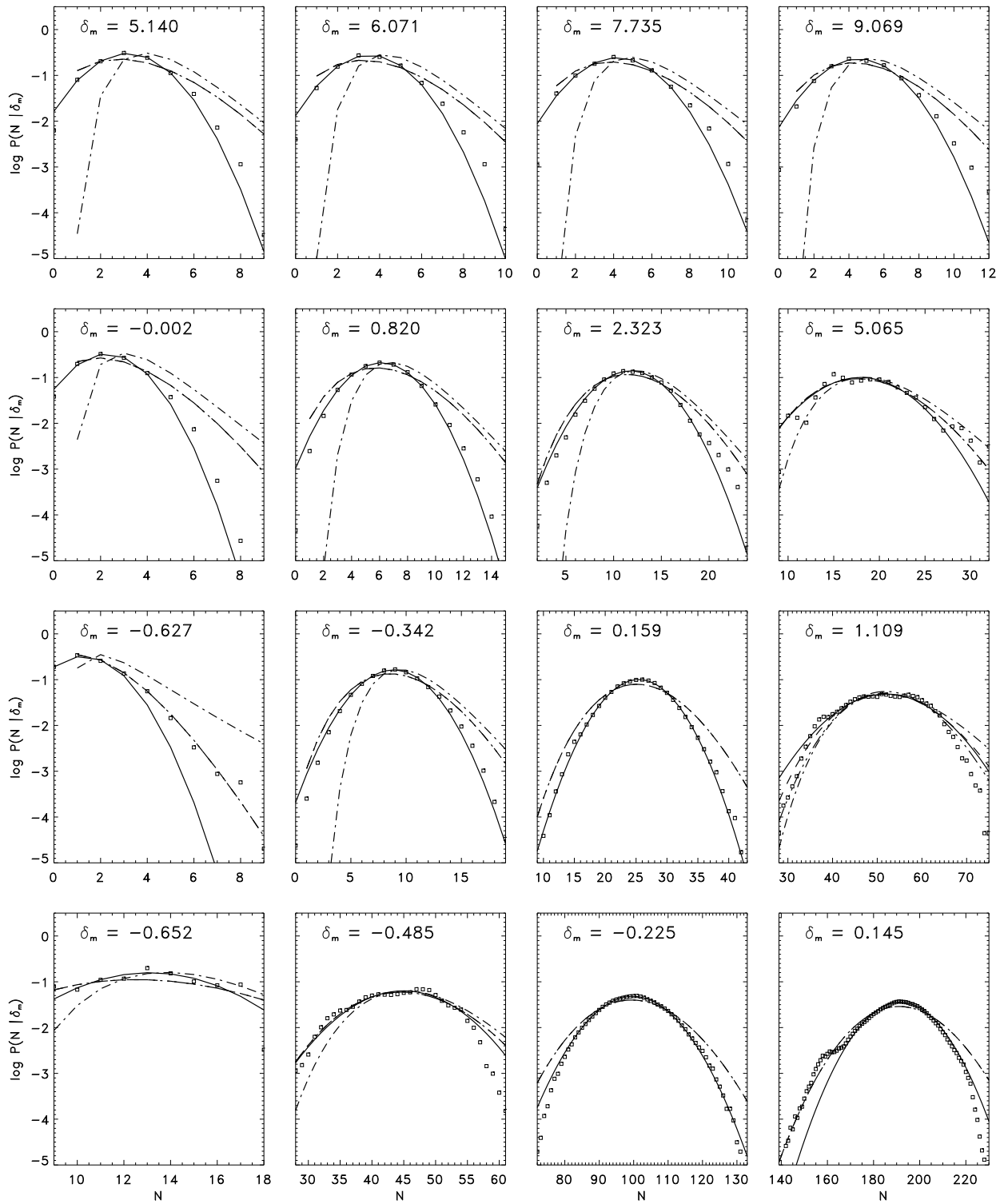
Figure 3.21: Comparison between the conditional probability measured from the numerical data (squares) for model galaxies at redshift 3 and the corresponding best fits of the Poisson (dash-dot-dot-dot line), Thermodynamic (dashed line), Lognormal (dash-dot line) and Gaussian (solid line) distribution functions. The rows correspond, from top to bottom, to the sampling scales 8.8, 17.6, 35.2 Mpc/h, respectively. For each sampling scale there are four plots corresponding to the local mass overdensity as indicated in the labels.
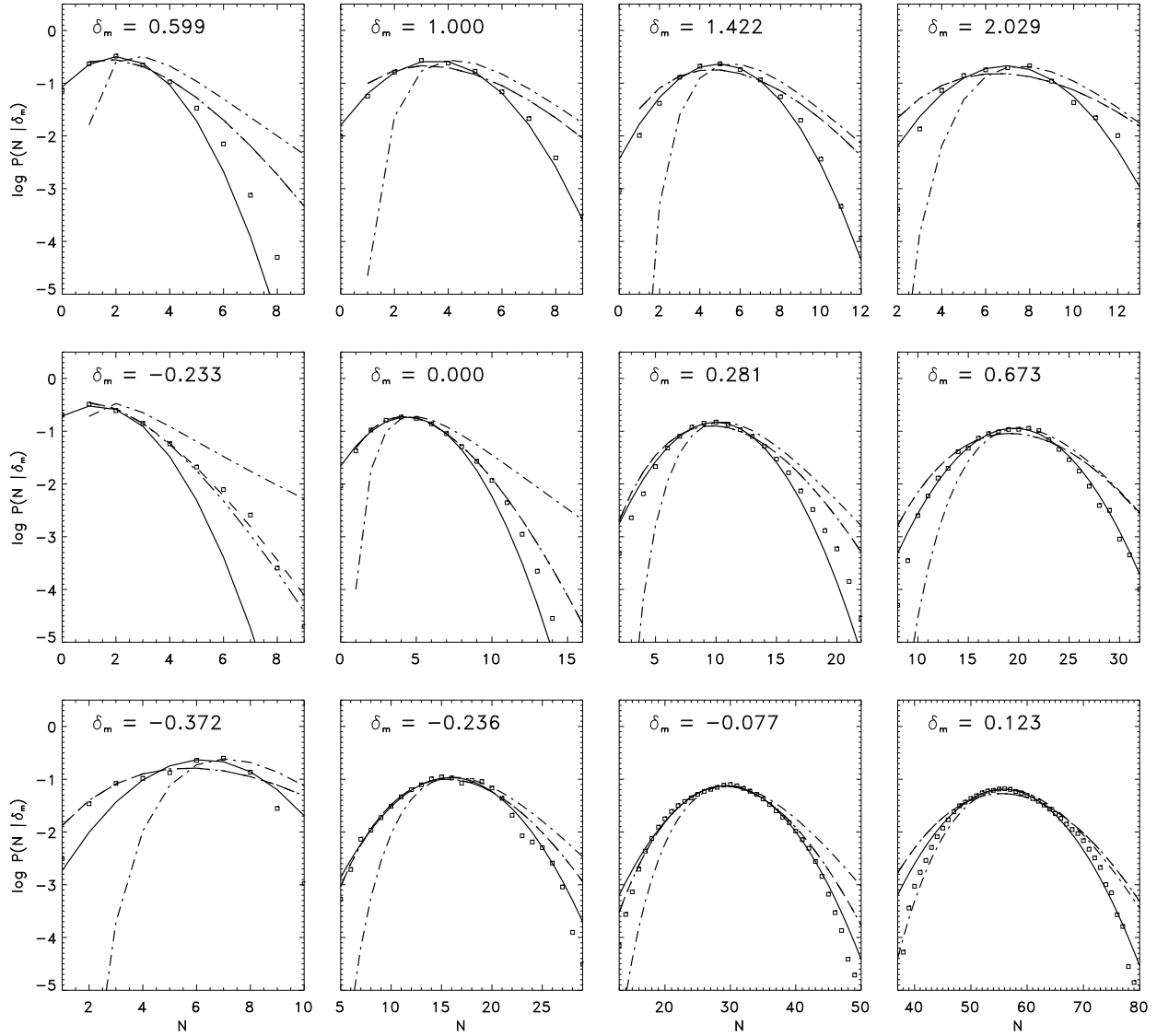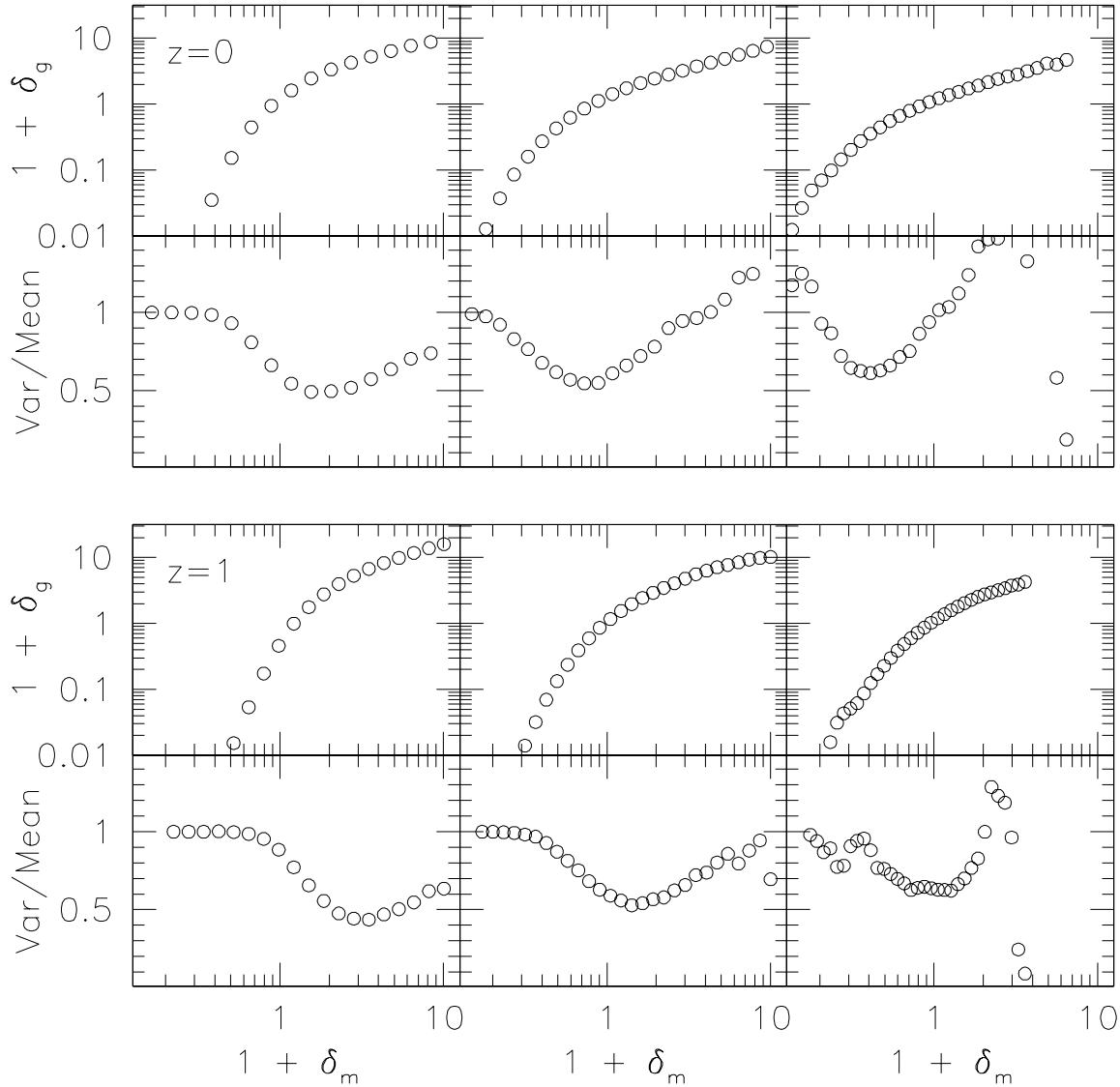
Figure 3.22: Mean bias relation and ratio between the variance and the mean of the bias relation of galaxies obtained from the simulations using semi-analytical models of galaxy formation. We show model galaxies at the present epoch (upper panel) and at redshift 1 (lower panel). The mean and ratio between the variance and the mean of the bias relation are shown in the top and bottom rows in each panel, respectively. At each epoch the cubical cells of side length, from left to right, $l = 4.4,\ 8.8,\ 17.6\ h^{-1}\ Mpc$ are shown.

3.4 Summary and Discussion

In this chapter the stochastic nature of the halo-mass bias relation has been investigated. In order to accomplish this purpose the conditional probability function $P_V(N|\delta_m)$ of haloes and mass, obtained from high resolution N-body simulations, has been analyzed in detail.

It has been found that the halo-mass bias relation from the simulations is well represented by a Gaussian model, and that a Poissonian model is generally a poor approximation to the numerical bias relation.

Furthermore, it has been shown that the model of Mo & White (1996) describes well the mean of the bias relation obtained from the simulations, confirming earlier results. The proposed extension to the model of Sheth & Lemson (1999) for the variance of the bias relation has been found to describe remarkably well the variance of the bias relation obtained from numerical simulations.

It has been shown that a simple phenomenological model for the halo-mass bias relation $P_V(N|\delta_m)$ can be constructed. The phenomenological model consists in describe the halo-mass bias relation by a Gaussian function with its mean as given by the model of Mo & White (1996) and the variance as given by the extension to the model of Sheth & Lemson (1999). This model allows one to construct a theoretical model for the full count-in-cell function for dark haloes.

In addition, the stochastic nature of the galaxy-mass bias relation has been investigated. It has been found that the galaxy distribution in the cosmic density field predicted by semi-analytic models of galaxy formation shows similar stochastic behavior to that of the haloes, implying that the galaxy distribution is not a Poisson sampling of the underlying density field. It has been found also that the conditional probability for galaxies is better described by a Gaussian function.

These results have important implications in the interpretations of galaxy clustering in terms of the underlying density field. For example, the quantity conventionally used to characterize the second moment of counts-in-cells is defined (here for dark halo) as

$$\kappa_2(R) = \frac{\langle (N - \overline{n}V)^2 \rangle}{(\overline{n}V)^2} - \frac{1}{(\overline{n}V)}, \tag{3.20}$$

where the second term on the right-hand side is to subtract Poisson shot noise (e.g. Peebles

1980). With the use of equation (3.1), it is easy to show that

$$
\begin{aligned}
\kappa_2(R) &= \frac{1}{(\bar{n}V)^2} \int \langle N|\delta_m\rangle^2 P_V(\delta_m)\, d\delta_m \\
&+ \frac{1}{(\bar{n}V)^2} \int \left[\sigma^2 - \langle N|\delta_m\rangle\right] P_V(\delta_m)\, d\delta_m - 1\,.
\end{aligned}
\tag{3.21}
$$

Thus, even if haloes trace mass on average, i.e. $\langle N|\delta_m\rangle \propto \delta_m$, this quantity is not equal to the second moment for the mass, because the second term on the right-hand side is generally non-zero. Thus, in order to infer the properties of the mass distribution in the Universe from statistical measures of the galaxy distribution, it is necessary to understand the stochastic nature of galaxy biasing.

Furthermore, the non-Poissonian behavior of the bias relation might imply that the (Poisson) shot-noise corrections usually applied at estimating higher-order moments of the galaxy distribution are not completely correct and therefore interpretations of skewness and kurtosis might change considerably, at least at the scales where shot-noise terms are not too small. This issue needs to be investigated in more detail. Thus, in order to infer the properties of the mass distribution in the universe from statistical measures of the galaxy distribution, it is necessary to understand first the stochastic nature of the galaxy-mass bias relation.

As discussed in Dekel & Lahav (1999), the stochasticity in galaxy biasing not only affects the interpretation of the moments of the galaxy distribution, but also affects the interpretation of other quantities related to statistical measures of galaxy clustering, such as, redshift distortions, the cosmic virial theorem and the cosmic energy equation.

Redshift Distortions

Comoving volume elements in redshift space are distorted in comparison to the corresponding volume elements in real space and thus, a large-scale isotropic distribution of galaxies in real space is observed as an anisotropic distribution in redshift space. These redshift distortions are caused by peculiar velocities along the line of sight. A very promising way of estimating $\beta \equiv \Omega^{0.6}/b$ is via redshift distortions in a redshift survey (Dekel & Lahav 1999, and references therein). The relation between peculiar velocities and the mass density depends on the value of Ω and, therefore, the distortions relative to the galaxy density depend on Ω and on the galaxy bias relation.

With their formalism for stochastic biasing, assuming no velocity biasing, and from the linear theory of gravitational instability Dekel & Lahav (1999) obtain a general local expression for redshift distortions

$$\sigma_{g,s}^2 = \sigma_g^2 [1 + 2(f(\Omega)\mu^2)rb_{\text{var}}^{-1} + ((f(\Omega)\mu^2)^2 \ b_{\text{var}}^{-2}],$$

where the subscript s denotes quantities measured in redshift space. $f(\Omega) \sim \Omega^{0.6}$ (Peebles 1980), μ^2 is a geometrical factor depending on the angle between the peculiar velocity (**v**) of the galaxy and its position in space (**x**). r and b_{var} are parameters describing the nonlinearity and stochasticity of the bias relation.

Cosmic Virial Theorem

The cosmic virial theorem can be used to estimate Ω from galaxy surveys (Peebles 1980; Bartlett & Blanchard 1996) by relating the observed dispersion of galaxy-galaxy peculiar velocities to a spatial integral over the three point galaxy-galaxy-mass cross correlation function (ξ_{ggm}) divided by the galaxy-galaxy correlation function (ξ_{gg}). The observable in this case is the three point galaxy correlation function. Therefore the corresponding biasing parameter is $b_{CV} = \langle \delta_g \ \delta_g \delta_g \rangle / \langle \delta_g \ \delta_g \delta_m \rangle$. At zero lag, the expression obtained by Dekel & Lahav (1999) is

$$b_{\text{CV}} = \frac{\langle \delta_m^3 \ b^3(\delta_m) \rangle + 3\sigma_m^2 \langle \delta_m b(\delta_m)\sigma_b^2(\delta_m) \rangle + S_b S_m}{\langle \delta_m^3 \ b^2(\delta_m) \rangle + \sigma_m^2 \langle \delta_m \sigma_b^2(\delta_m) \rangle},$$

where $b(\delta_m)$ is the mean biasing function, $\sigma_b^2(\delta_m)$ is the biasing scatter function, S_b is the biasing skewness function and S is the third moment of the mass distribution. Recall that our model assumes a Gaussian conditional probability function, so $S_b \equiv 0$.

Cosmic Energy Equation

The cosmic energy equation (Peebles 1980) 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 ($\xi_{gm}(r)$). In this case the observable quantity corresponds to the galaxy-galaxy correlation function ($\xi_{gm}(r)$), therefore, the necessary biasing parameter is $b_{CE} = \langle \delta_g \ \delta_g \rangle / \langle \delta_g \ \delta_m \rangle$. At zero lag, this biasing parameter takes the form of $b_{\text{CE}} = b_{\text{inv}} = b_{\text{var}}/r$ in the formalism of Dekel & Lahav (1999).

With the results found in this investigation, one might be able to model some of these effects quantitatively.

Chapter 4

Concluding Remarks

4.1 Results and Conclusions

In this thesis I have studied the bias relation between the spatial distribution of dark matter haloes and the spatial matter density field. The theoretical background for the formation of structure in the universe corresponds to the gravitational collapse of initially Gaussian density fluctuations in cold dark matter universes, which is currently the best description of the process of structure formation in the universe.

The construction of feasible theoretical models of structure formation requires the knowledge of all the physical processes involved in the formation and clustering of galaxies. In addition the models must reproduce the clustering patterns observed from galaxy catalogues. The most challenging problem in the construction of these models corresponds to the understanding of the relationship between the galaxy distribution and the underlying matter distribution in the universe. From earlier studies it is known that, in general, galaxies do not trace the underlying mass exactly and that if the structure has evolved according to standard gravitational instability theory then the galaxy distribution must be biased respect to the total mass distribution. All these facts together mean that it is mandatory to understand the process of galaxy biasing if one wants to constrain models of galaxy formation or to constrain the values of cosmological parameters from the observed distribution of galaxies.

Initially proposed by White & Rees (1978), the current framework of galaxy formation is divided into two parts: first, the dominant dark matter component in the universe

collapses by gravitational instability into small lumps which then undergo a hierarchical process of formation of larger structures; second, the gas fraction trapped within the potential wells of the dark matter lumps cools down and condenses to form galaxies. While the first stage of galaxy formation is easy to solve, since it is driven by gravity alone, the second stage is not yet well understood and many physical processes are involved in it. Nevertheless, these processes are known to have their main influence on the individual properties of galaxies but a negligible influence on the overall clustering properties. That means that the clustering properties of galaxies are mainly determined by the gravitational processes involved in the formation and clustering of dark matter haloes (i.e. virialized dark matter clumps) and therefore, the clustering properties of the galaxy distribution can be fairly approached by studying the clustering properties of the dark matter haloes.

In this thesis I have studied the clustering properties of the spatial distribution of dark matter haloes in cosmic density fields. The analysis has been divided into two main parts. The first part corresponds to the study of deterministic halo-mass bias models, based on the spherical collapse model as well as on the ellipsoidal collapse model. The second part corresponds to the study of the stochasticity in the halo-mass bias relation.

In Chapter 1 the analysis of the deterministic halo-mass bias has been presented. There I have tested the spherical collapse models from Mo & White (1996) for the variance and the model from Mo et al. (1997) for the higher-order moments of the halo counts-in-cells, two sets of high-resolution N-body simulations with different simulation boxes and mass resolution. Furthermore, the extensions of these models, based on the ellipsoidal collapse (Sheth et al. 2001), have also been tested.

From the set with a very large simulation box and low mass resolution (VIRGO Simulations), which allows one to control the finite volume effect, it has been found that the biasing models under analysis work remarkably good for massive haloes in cold dark matter universes. The good performance of the biasing models when the moments from the mass distribution are estimated using the linear perturbation theory, shows that the moments from this simulations set are practically unaffected by the finite volume effect. The other set of simulations (GIF simulations), which has much higher mass resolution and smaller box-size, has been used to test the biasing models for low-mass haloes, I have shown that a significant improvement can be achieved for haloes less massive than $M^\star$ if the ellipsoidal collapse model is used instead of the spherical collapse model in defin-

ing dark haloes. For massive haloes both the Mo & White (1996) and Mo et al. (1997) biasing models and their ellipsoidal extension (Sheth et al. 2001) work remarkably good.

The theoretical models have been used to predict the high-order moments at a fixed scale of the Lyman break galaxies observed at $z = 3$ and their descendants at lower redshifts, which are commonly assumed to form in the center of the most massive haloes at redshift ~ 3 (Mo & Fukugita 1996; Adelberger et al. 1998; Jing & Suto 1998; Mo et al. 1999). Under this assumption and, supposing that only a negligible fraction of those haloes host a secondary observable galaxy the observed Lyman Break Galaxies at redshift 3 correspond to the most massive haloes at $z = 3$. It has been found that, although the linear bias parameter b depends strongly on the cosmology adopted, the values of the high-order moments are practically the same in both ΛCDM and τCDM dark matter universes and therefore the high-order moments from the spatial distribution of these objects cannot constrain cosmological parameters.

In chapter 2 the stochastic nature of the halo-mass bias relation is investigated. In order to accomplish this purpose the conditional probability functions $P_V(N|\delta_m)$ of haloes and mass, obtained from high resolution N-body simulations, have been analyzed in detail.

It has been found that the halo-mass bias relation from the simulations is well represented by a Gaussian model, and that the commonly adopted Poissonian model is, in general, a poor approximation to the numerical halo-bias relation. That means that the galaxy biasing process, as well as the halo biasing process, is not only determined by the local value of the mass density field, but also by other local quantities, such as clumpiness, and by non-local properties, such as large-scale tidal fields.

Furthermore, it has been shown that the model of Mo & White (1996) describes well the mean of the bias relation obtained from the simulations, confirming earlier results. The proposed extension to the model of Sheth & Lemson (1999) for the variance of the bias relation has been found to describe remarkably well the variance of the bias relation obtained from numerical simulations.

It has been shown, additionally, that a simple phenomenological model for the halo-mass bias relation $P_V(N|\delta_m)$ can be constructed. The phenomenological model consists in describing the halo-mass bias relation by a Gaussian function with its mean as given by the model of Mo & White (1996) and the variance as given by the extension to the model of Sheth & Lemson (1999). This model might allow one to construct a theoretical

model for the full count-in-cell function for dark haloes.

In addition, the stochastic nature of the galaxy-mass bias relation has been investigated. It has been found that the galaxy distribution in the cosmic density field predicted by semi-analytic models of galaxy formation (Kauffmann et al. 1999) shows similar stochastic behavior to that of the haloes, implying that the galaxy distribution is not a Poisson sampling of the underlying density field. It has been found also that the conditional probability for galaxies is well described by a Gaussian function.

These results have important implications in the interpretations of galaxy clustering in terms of the underlying density field. For example, the non-Poissonian behavior of the bias relation might imply that the (Poisson) shot-noise corrections usually applied at estimating higher-order moments of the galaxy distribution are not completely correct and therefore interpretations of skewness and kurtosis might change considerably, at least at the scales where shot-noise terms are not too small. This issue needs to be investigated in more detail. Thus, in order to infer the properties of the mass distribution in the Universe from statistical measures of the galaxy distribution, it is necessary to understand first the stochastic nature of galaxy biasing.

The stochasticity in galaxy biasing not only affects the interpretation of the moments of the galaxy distribution, but also affects the interpretation of other quantities related to statistical measures of galaxy clustering, such as, redshift distortions, the cosmic virial theorem and the cosmic energy equation. With the results shown in chapter 2 one might be able to model quantitatively many of these effects.

4.2 Future Prospects

The several physical processes involved in the formation of galaxies are currently poorly understood. However, our understanding of the universe, and the formation of galaxies within it, will be substantially improved in the next years thanks to forthcoming new observational data and the theoretical progress linked to it. Indeed, for some years now the *Hubble Space Telescope* has made possible the exploration of the universe at high redshift. In addition to that, the new huge telescopes like *VLT* and *Keck* will give us an increasingly better picture of the universe at the epochs of the birth of galaxies, and therefore will help us to improve considerably the framework of galaxy formation. In

addition, the new *Sloan Digital Sky Survey (SDSS)* and *2dF* huge redshift surveys are going to produce extended catalogues of very high quality.

On the other hand, precise measurements of the cosmological parameters could be obtained from the *Planck-Surveyor* and *MAP*, if they are successful.

The high-quality observational data represent a theoretical challenge, in the sense that we will need to improve our models of structure formation and clustering to be consistent with the observations. We also need to improve numerical simulations, in order to be able to reproduce the observations from semi-analytical models, as well as to be able to work out and improve models of galaxy and halo biasing, which is mandatory if one wants to use galaxy clustering measurements to understand the mass distribution in the universe and to constrain the values of cosmological parameters. Thus, this represents a huge opportunity to go forwards in our understanding of the universe.

Bibliography

Adelberger K. L., Steidel C. C., Giavalisco M., Dickinson M. E., Pettini M., Kellog M., 1998, ApJ, 505, 18

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Bartlett J. G., Blanchard A., 1996, Astron. Astrophys., 307, 1

Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, MNRAS, 320, 261

Bernardeau F., 1992, ApJ, 392, 1

Bernardeau F., 1994, ApJ, 433, 1

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bond J. R., Myers S. T., 1996, ApJS, 103, 1+

Casas-Miranda R., Mo H. J., Boerner G., 2002, MNRAS (submitted)

Casas-Miranda R., Mo H. J., Sheth R. K., Boerner G., 2002, MNRAS, 333, 730

Colberg J. M., White S. D. M., Yoshida N., MacFarland T. J., Jenkins A., Frenk C. S., Pearce F. R., Evrard A. E., Couchman H. M. P., Efstathiou G., Peacock J. A., Thomas P. A., The Virgo Consortium 2000, MNRAS, 319, 209

Cole S., Kaiser N., 1989, MNRAS, 237, 1127

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Coles P., 1993, MNRAS, 262, 1065

Coles P., Jones B., 1991, MNRAS, 248, 1

Coles P., Lucchin F., 1995, Cosmology. The origin and evolution of cosmic structure. Chichester: Wiley, —c1995

Colombi S., Bouchet F. R., Shaeffer R., 1994, Astron. Astrophys., 281, 301

Dekel A., Lahav O., 1999, ApJ, 520, 24

Dressler A., 1980, ApJ, 236, 351

Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P

Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715

Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447

Governato F., Babul A., Quinn T., Tozzi P., Baugh C. M., Katz N., Lake G., 1999, MNRAS, 307, 949

Jenkins A., Frenk C. S., Pearce F. R., Thomas P. A., Colberg J. M., White S. D. M., Couchman H. M. P., Peacock J. A., Efstathiou G., Nelson A. H., 1998, ApJ, 499, 20

Jing Y. P., 1998, ApJL, 503, L9

Jing Y. P., Mo H. J., Boerner G., 1998, ApJ, 494, 1

Jing Y. P., Suto Y., 1998, ApJ, 494, L5

Kaiser N., 1984, ApJL, 284, L9

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188

Lacey C., Cole S., 1994, MNRAS, 271, 676

Ma C. P., Fry J. N., 2000, ApJ, 531, 87

Mo H., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189

Mo H. J., Fukugita M., 1996, ApJL, 467, L9

Mo H. J., Jing Y. P., Börner G., 1997, MNRAS, 286, 979

Mo H. J., Jing Y. P., White S. D. M., 1996, MNRAS, 282, 1096

Mo H. J., Mao S., White S. D. M., 1999, MNRAS, 304, 175

Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

Peacock J. A., 1999, Cosmological physics. Cambridge University Press, Cambridge, UK.

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press

Press W. H., Schechter P., 1974, ApJ, 187, 425

Saslaw W. C., Hamilton A., 1984, ApJ, 276, 13

Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20

Seljak U., 2000, MNRAS, 318, 203

Sheth R. K., 1998, MNRAS, 300, 1057

Sheth R. K., Lemson G., 1999, MNRAS, 304, 767

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1+

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Somerville R. S., Lemson G., Sigad Y., Dekel A., Kauffmann G., White S. D. M., 2001,
    MNRAS, 320, 289

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Szapudi I. ., Quinn T., Stadel J., Lake G., 1999, ApJ, 517, 54

Totsuji H., Kihara T., 1969, PASJ, 21, 221+

White S. D. M., Frenk C., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Acknowledgements

I am very grateful to Professor Gerhard Börner and Dr. Houjun Mo for their invaluable advice and for their encouragement in my research work. I also thank the Max-Planck-Institut für Astrophysik for the hospitality and research facilities it offered me during my graduate studies.

Finally I want to acknowledge the financial support from the "Francisco José de Caldas Institute for the Development of Science and Technology (COLCIENCIAS)".

Lebenslauf

Persönliche Daten

Rigoberto Angel Casas Miranda

geboren am 18. März 1971 in Bogotá, Kolumbien

Verheiratet mit Astrid Baquero Bernal

Schulausbildung

1978 - 1982   Grundschule in Bogotá, Kolumbien

1983 - 1989   Sekundarschule in Bogotá, Kolumbien

Wehrdienst

Jan. - Dez. 1989   Wehrmacht Kolumbiens

Studium

1990-1995   Studium der Physik, Universidad Nacional de Colombia, Bogotá

1995-1997   Master in Naturwissenschaften, Universidad Nacional
              de Colombia, Bogotá

1997 - 2001   Doktorand am Max-Planck-Institut für Astrophysik, München

Auszeichnungen

1997-2001   Kolumbianisches Institut für Wissenschaft und Technische Entwicklung
            "FRANCISCO JOSE DE CALDAS" (COLCIENCIAS) Stipendium
            um eine Promotion im Ausland zu absolvieren, Kolumbien

1996-1997   COLCIENCIAS Stipendium für junge WissenschaftlerInnen, Kolumbien

1996   MAZDA-Stiftung Stipendium, Kolumbien

1994   Ausgewählt im WissenschaflterInnen-Programm der Wissenschaft,
            Ausbildung und Entwicklung Mission. Präsidentschaft der Republik von
            Kolumbien, Beratungsamt für die Modernisierung des Staates und
            COLCIENCIAS, Kolumbien.