Galaxy Bias from the Bispectrum: A New Method

Jennifer E. Pollack
University of Bonn

Thesis Supervisor: Cristiano Porciani
Collaborator: Robert E. Smith

Motivation

1) Main Objective of state-of-the-art surveys such as Euclid is to map the geometry of the dark universe in order to test models of cosmology

→ One key is to probe the total matter density in the Universe, which is dominated by dark matter via weak lensing and galaxy clustering

2) However, galaxies as well as their host dark matter haloes are known to be biased tracers of the underlying total matter distribution.

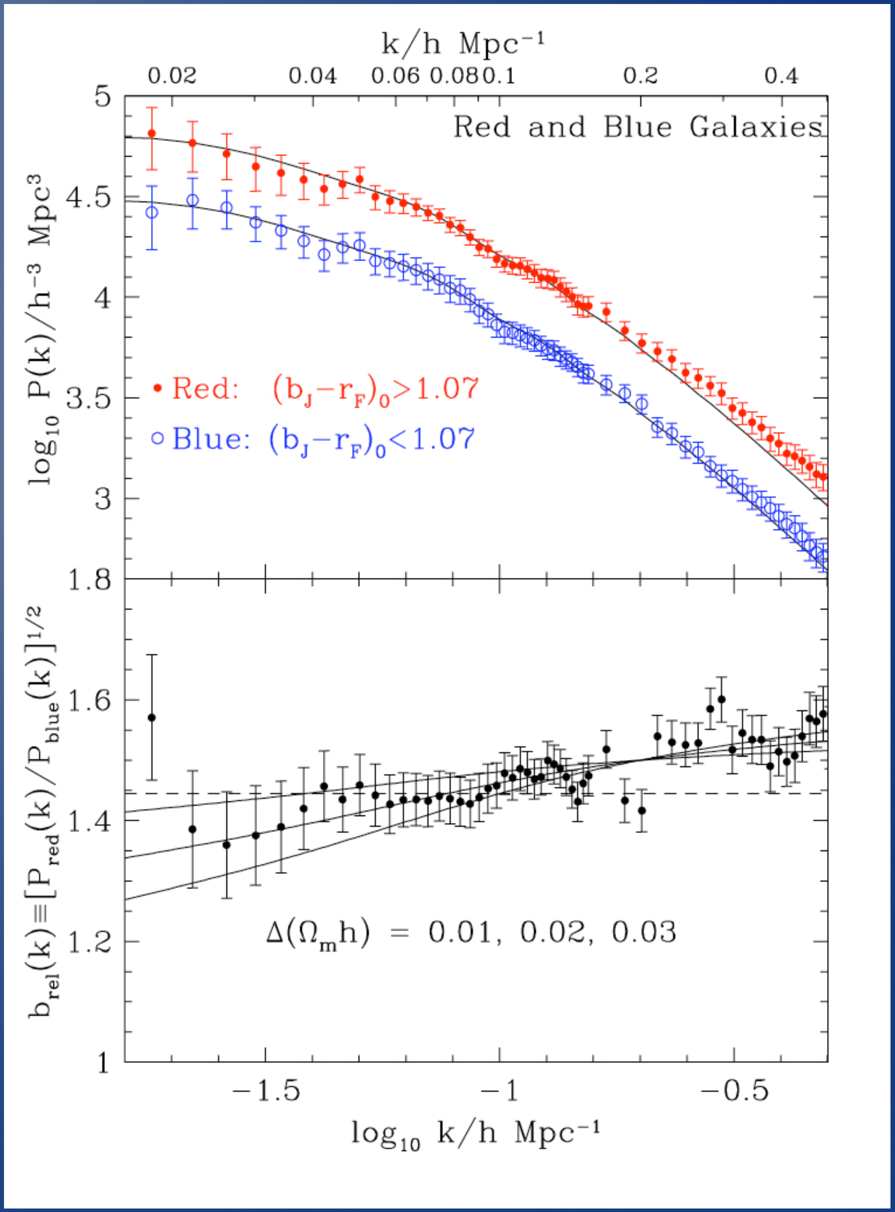
3) the bispectrum, Fourier analogue to 3-point correlation function is a useful statistical probe into galaxy/halo bias

4) Cosmological n-body simulations are an invaluable tool for testing models of structure growth, biasing relations, etc.

→ Beyond model-testing, they may have more application in model building.

Galaxy bias

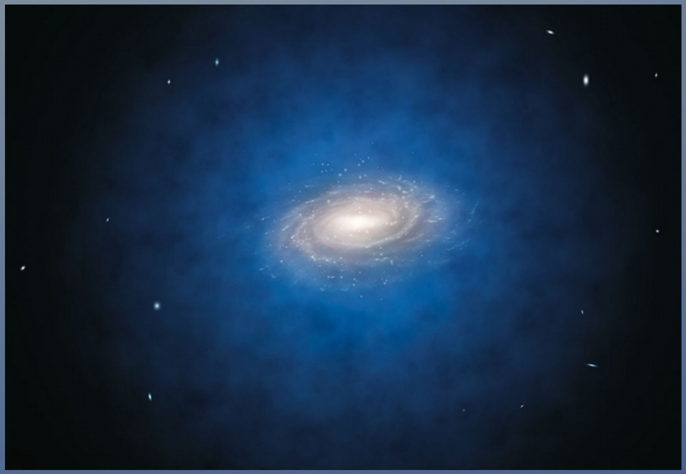
- Amplitude and shape of power spectrum of different tracer populations of galaxies varies with luminosity, color, and morphology.
- Red galaxies more clustered than blue galaxies and the level of clustering increases with progression to smaller scales
- Thus, not all galaxy types are unbiased tracers of the matter



Cole et al. 2005

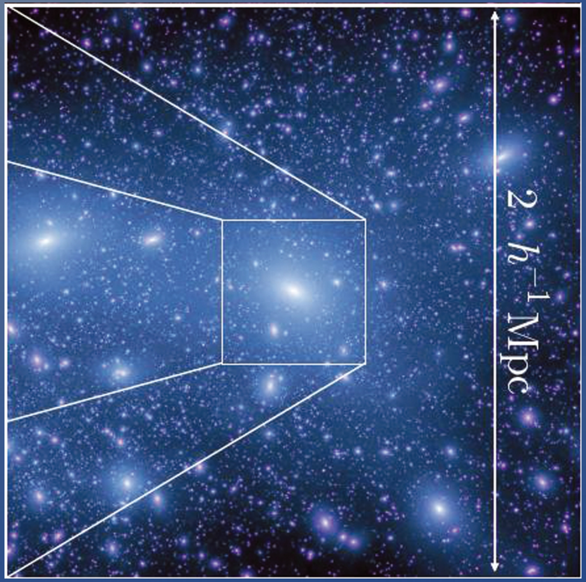
From Galaxy Bias to Halo Bias...

Galaxies are thought to form in the gravitational potential wells of collapsed dark matter halos.



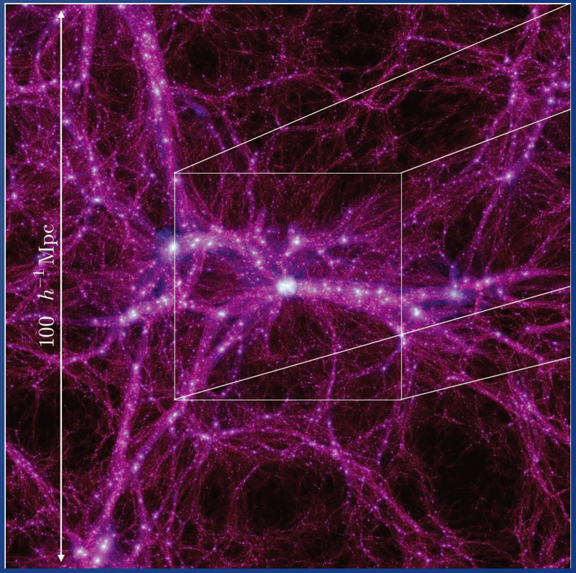
Credit: ESO/L. Calcada 2012

This motivates the need to understand how the clustering properties of dark matter haloes



Credit: Boylan-Kolchin et al. 2009

relates to the clustering of the total dark matter distribution.

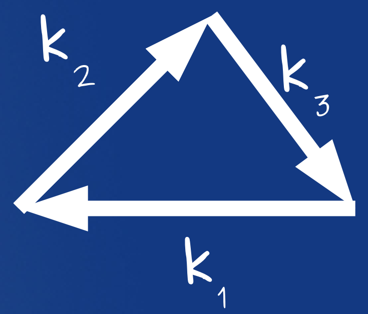


Credit: Boylan-Kolchin et al. 2009

Investigating Bias with the Bispectrum

Formally, the Bispectrum is defined as:

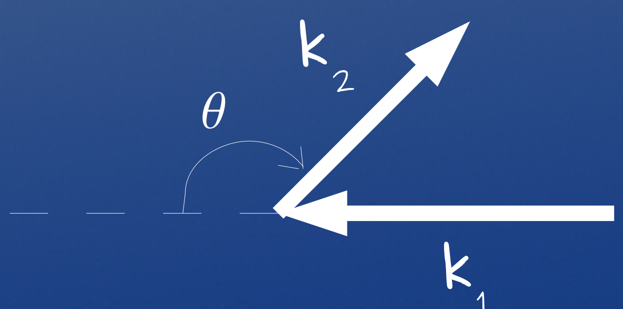
$$\langle\, \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\,\rangle \equiv (2\pi)^3\delta^{\mathrm{D}}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$$

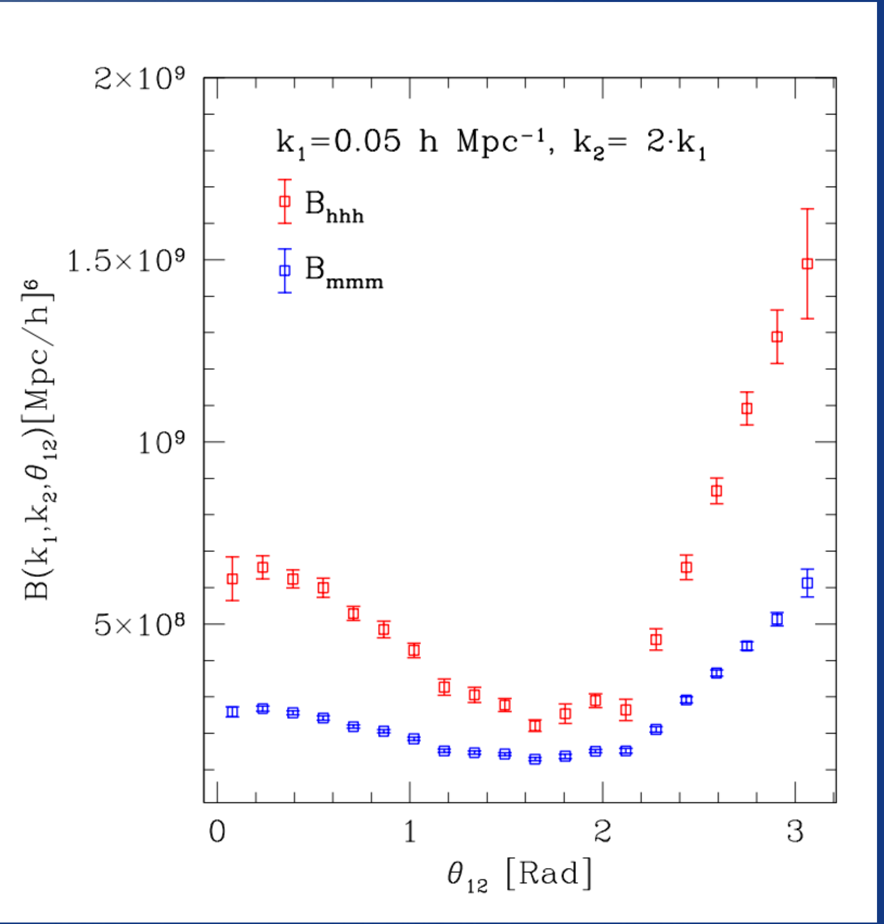


Bispectrum is regarded as the go-to statistic to constrain galaxy/halo bias

Why? Because the bispectrum carries a specific triangle configuration dependence that can be used to distinguish between different models of bias such as local, linear and non-linear bias.

Bispectrum can be measured for different sets of triangles with two side lengths kept fixed, and the angle between them is varied

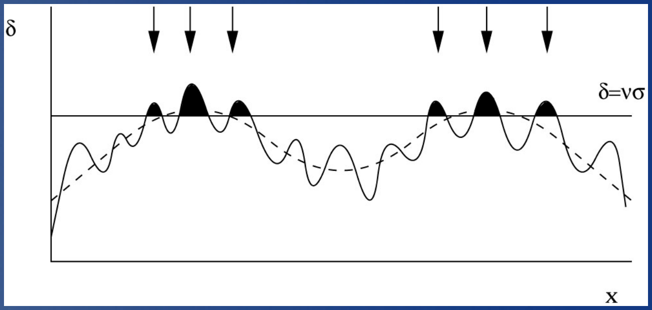




Modeling galaxy/halo bias

What can the galaxy/halo density field depend on?

It is assumed galaxies and their hosting dark matter halos are biased because their formation occurs at the high peaks of the density field filtered at small scales.



Types of modeling

1) local and deterministic bias (galaxies/haloes can be painted onto the mass distribution)

A) clustering of haloes at a given location only depends on the clustering of the total matter at the same location and time when the haloes were identified (Eulerian bias)

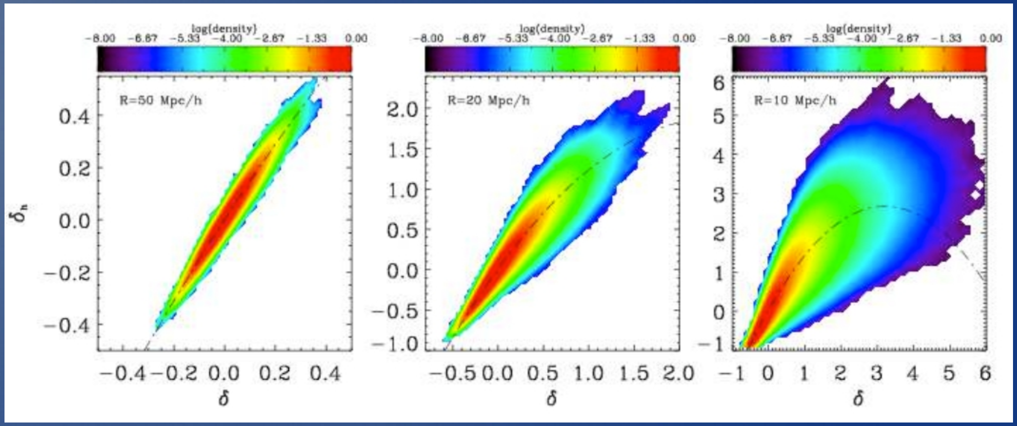
2) Non-local bias extensions (actual dynamics involved)

a) model environmental effects like tidal fields & galaxy/halo mergers

3) Stochastic component

A) stochastic effects induced by environmental effects during formation & evolution

B) galaxy/halo distribution is a point process generates shot-noise in statistics



Pollack, Smith, & Porciani 2012

Local Eulerian Biasing Overview

Let the halo density field, smoothed on scale R, be a function of the local matter density in the form of a Taylor Series Expansion about $\delta(R)=0$ (Fry & Gaztanaga 1993, Coles 1993)

$$\delta_{\mathrm{h}}(\boldsymbol{x}|M, R) = \sum_{j=0}^{\infty} \frac{b_j(M)}{j!}[\delta(\boldsymbol{x}|R)]^j$$

$$\delta(\mathbf{x}|R) = \frac{1}{V}\int d^3y\,\delta(\mathbf{y})W(|\mathbf{x}-\mathbf{y}|, R)$$

Taking the Fourier Transform

$$\delta_{\mathrm{h}}(\mathbf{k}|M,R) = b_1(M)\delta(\mathbf{k}|R) + \frac{b_2(M)}{2}\int\frac{\mathrm{d}^3\mathbf{q}_1}{(2\pi)^3}\delta(\mathbf{q}_1|R)\delta(\mathbf{k}-\mathbf{q}_1|R) + \ldots$$

where, $\delta_i(\mathbf{q}_j|R) \equiv W(|\mathbf{q}_j|R)\delta_i(\mathbf{q}_j)$

Typical application of local Eulerian bias in higher-order statistics

1) Employ standard perturbation theory (SPT) to model the non-linear matter density field whose evolution was induced by gravitational instability.

SPT solves fluid equations of motion (e.g. continuity, Euler, Poisson) and then applies a perturbative expansion about the linear solution, $\delta(k, t) = D1(t)\delta_1(k)$

Non-linear PT:

$$\delta(\mathbf{k},t) = \sum_{n=1}^{\infty} D_1(t)^n \delta_n(\mathbf{k})$$

$$D_1(t = \mathrm{today}) = 1$$

$$\delta_n(\boldsymbol{k}) = \int \frac{\mathrm{d}^3\boldsymbol{q}_1}{(2\pi)^3} \cdots \int \frac{\mathrm{d}^3\boldsymbol{q}_n}{(2\pi)^3} (2\pi)^3 \delta^{\mathrm{D}}(\boldsymbol{k} - \boldsymbol{q}_1 - \cdots - \boldsymbol{q}_n) \times F_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \delta_1(\boldsymbol{q}_1) \cdots \delta_1(\boldsymbol{q}_n) \,.$$

2) Insert SPT expansion of $\delta(k)$ (smoothed) up to some order (usually second-order) into model for the halo density field

$$\delta_{\mathrm{h}}(\mathbf{k}|M,R) = b_1(M)[\delta_1(\mathbf{k}|R) + \delta_2(\mathbf{k}|R)] + \frac{b_2(M)}{2} \int \frac{\mathrm{d}^3\mathbf{q}_1}{(2\pi)^3} \delta_1(\mathbf{q}_1|R)\delta_1(\mathbf{k}-\mathbf{q}_1|R)$$

3) Using this approach one finds a perturbative expansion of halo power- & bi-spectra

$$\begin{aligned} P_{\mathrm{hh}}(M) &= P_{\mathrm{hh}}^{(0)}(M) + P_{\mathrm{hh}}^{(1)}(M) + \ldots \\ B_{\mathrm{hhh}}(M) &= B_{\mathrm{hhh}}^{(0)}(M) + B_{\mathrm{hhh}}^{(1)}(M) + \ldots \end{aligned}$$

Typical application of local Eulerian bias in higher-order statistics

4) Consider lowest-order non-vanishing terms, so called tree-level terms, which are of the form:

$$
\begin{aligned}
P^{(0)}_{\rm hh}(k|M,R) &= b_1^2(M)P^{(0)}_{\rm mm}(k|R) \\
B^{(0)}_{\rm hhh}(\mathbf{k}_1,\mathbf{k}_2|M,R) &= b_1^3(M)B^{(0)}_{\rm mmm}(\mathbf{k}_1,\mathbf{k}_2|R) + b_1^2(M)b_2(M)\left[P^{(0)}_{\rm mm}(k_1|R)P^{(0)}_{\rm mm}(k_2|R) + 2\,{\rm cyc}\right]
\end{aligned}
$$

5) Smoothing is problematic. We do not smooth the halo density field apart from the mass assignment scheme used to compute the density fields. And, in actuality for galaxy density fields we do not know a priori what the smoothing scale is.

So, one way out is to Conjecture: 'de-smooth' (Smith et al 2007, 2008)

$$
\begin{aligned}
P^{(0)}_{\rm hh}(k|M) &= \frac{P^{(0)}_{\rm hh}(k|M,R)}{W^2(kR)} \; ; \\
B^{(0)}_{\rm hhh}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3|M) &= \frac{B^{(0)}_{\rm hhh}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3|M,R)}{W(\mathbf{k}_1R)W(\mathbf{k}_2R)W(\mathbf{k}_3R)}
\end{aligned}
$$

6) Applying de-smoothing almost fixes things

In limit of very large scales or arbitrarily small smoothing scales, $k_i R \to 0$

$$
B^{(0)}_{\rm hhh}(\mathbf{k}_1,\mathbf{k}_2|M) = b_1^3(M)B^{(0)}_{\rm mmm}(\mathbf{k}_1,\mathbf{k}_2) + b_1^2(M)b_2(M)\frac{W(|\mathbf{k}_1|R)W(|\mathbf{k}_2|R)}{W(|\mathbf{k}_3|R)}\left[P^{(0)}_{\rm mm}(k_1)P^{(0)}_{\rm mm}(k_2) + 2\,{\rm cyc}\right].
$$

Typical application of local eulerian bias in higher-order statistics

7) Lastly, at the three-point level, galaxy bias has often been modelled in terms of the lowest order hierarchical amplitude (predicted by SPT), the so-called reduced bispectrum, Q.

Reduced Bispectrum:

$$Q(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{P(k_1)P(k_2) + \text{cyc}}$$

Tree-level Halo Reduced Bispectrum:

$$Q^{(0)}_{\text{hhh}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3|M) = \frac{Q^{(0)}_{\text{mmm}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3|M)}{b_1(M)} + \frac{b_2(M)}{b_1^2(M)}$$

Favored statistic since it's almost close to unity and at tree-level it's insensitive to time evolution and cosmology

N-Body Simulations

Zbox2 photo credit of ITP-UZH

125 quad opteron 852 nodes and a total of 580Gb RAM and 65 Tb of disk. The nodes are connected with a high speed 3D SCI network.

We analyze 200* cosmological dark matter n-body simulations on zbox-2 and zbox-3 supercomputers at the university of zurich

Comoving Box size: 1.5 Gpc/h

No. of Particles=750^3

Cosmological model simulated

- Gaussian initial conditions set at z=49 using 2LPT (Crocce et al. 2006)
- Lambda cold dark matter model (WMAP Komatsu et al. 2009)

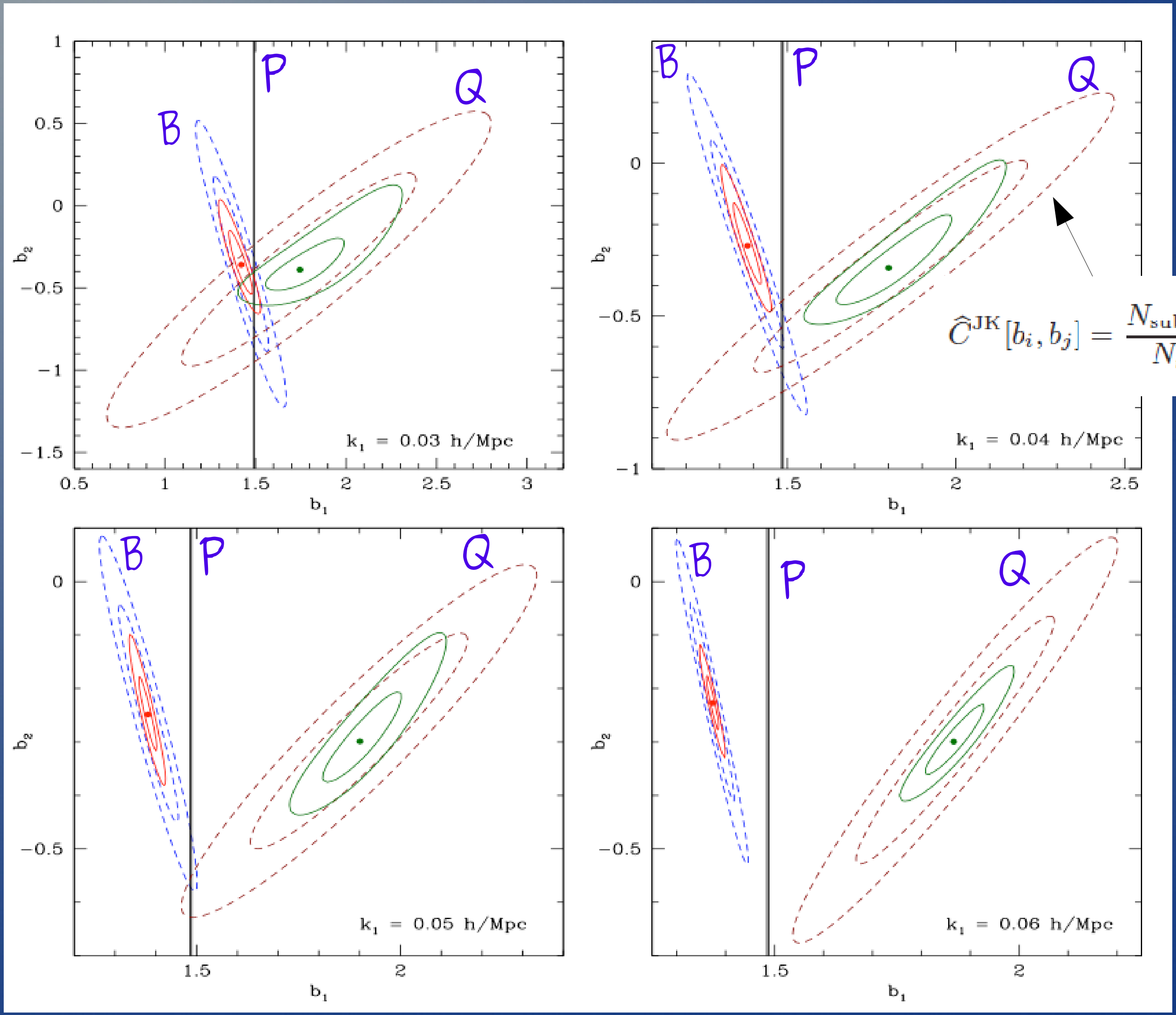
Ω_M=0.25, Ω_Λ=0.75,σ_8=0.8, n=1,h=0.7

Dark matter halo catalogues

- Friends-of-friends haloes
- Number of Haloes = 1.26×10^6
- Minimum Mass = $1.11 \times 10^{13}\ M_\odot/h$

*Results presented in this talk is for a subsample of 40 simulations

Performance of the tree-level bias models* for different scale ranges from N-body simulations



$$\widehat{C}^{\rm JK}[b_i, b_j] = \frac{N_{\rm sub}-1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} (b_{i,k} - \hat{b}_j)(b_{j,k} - \hat{b}_j)$$

*In comparison with effective bias from Power spectrum ratios taken strictly from simulations

$$b_{\rm hh}^{\rm eff} = \frac{1}{N_{\rm s}} \sum_{i=1}^{N_{\rm s}} \sqrt{\frac{\widehat{P}_{\rm hh}(k_i)}{\widehat{P}_{\rm mm}(k_i)}}$$

Pollack, Smith, Porciani 2012

Beyond tree-level bias models: 'biasing by hand'

Goal: Assume local bias model at second-order is the true model, determine how well the tree-level theory can work

'Bias-by-hand'
The smoothed matter density

$$\delta_{\rm h}(\mathbf{x}|R) = b_1\delta(\mathbf{x}|R) + \frac{b_2}{2}\delta^2(\mathbf{x}|R)$$

$$\delta_{\rm h}(\mathbf{k}|M,R) = b_1(M)\delta(\mathbf{k}|R) + \frac{b_2(M)}{2}\int\frac{{\rm d}^3\mathbf{q}_1}{(2\pi)^3}\delta(\mathbf{q}_1|R)\delta(\mathbf{k}-\mathbf{q}_1|R) + \ldots$$

From N-body simulations, measure ensemble-averaged B_{HHH}, B_{HHM}, B_{HMM}

Auto- and cross-bispectra are given exactly by sums of 3-, 4-, 5-, 6-point functions

$$\begin{aligned}
\mathcal{B}_{\rm hmm} &= b_1\mathcal{B}_{\rm mmm} + \frac{b_2}{6}\mathcal{P}_{4,\rm m}\ ;\\
\mathcal{B}_{\rm hhm} &= b_1^2\mathcal{B}_{\rm mmm} + \frac{b_1 b_2}{3}\mathcal{P}_{4,\rm m} + \frac{b_2^2}{12}\mathcal{P}_{5,\rm m}\ ;\\
\mathcal{B}_{\rm hhh} &= b_1^3\mathcal{B}_{\rm mmm} + \frac{b_1^2 b_2}{2}\mathcal{P}_{4,\rm m} + \frac{b_1 b_2^2}{4}\mathcal{P}_{5,\rm m} + \frac{b_2^3}{8}\mathcal{P}_{6,\rm m}
\end{aligned}$$

Inverse solve to get P_4, P_5, P_6

Beyond tree-level bias models: 'biasing by hand'

I: Exact model: Perfect recovery of b1=1.63 & b2=-0.53

Pollack, Smith, Porciani 2012

II: Exact Trispectrum:

$R\,[h^{-1}\mathrm{Mpc}]$		$b_1 \pm \sigma_{b_1}$	$b_2 \pm \sigma_{b_2}$	χ^2
20	B_{bbb}	1.62 ± 0.07	-0.46 ± 0.12	0.01
20	B_{bbm}	1.62 ± 0.10	-0.49 ± 0.19	0.00
20	B_{bmm}	1.63 ± 0.22	-0.53 ± 0.51	0.00
10	B_{bbb}	1.62 ± 0.04	-0.42 ± 0.04	0.15
10	B_{bbm}	1.62 ± 0.06	-0.47 ± 0.07	0.02
10	B_{bmm}	1.63 ± 0.13	-0.53 ± 0.20	0.00
6.7	B_{bbb}	1.59 ± 0.04	-0.35 ± 0.02	0.70
6.7	B_{bbm}	1.60 ± 0.05	-0.42 ± 0.04	0.14
6.7	B_{bmm}	1.63 ± 0.12	-0.53 ± 0.12	0.00

III: Tree-level:

$R\,[h^{-1}\mathrm{Mpc}]$		$b_1 \pm \sigma_{b_1}$	$b_2 \pm \sigma_{b_2}$	χ^2
20	B_{bbb}	1.63 ± 0.11	-0.67 ± 0.36	14.75
20	B_{bbm}	1.58 ± 0.17	-0.45 ± 0.65	15.71
20	B_{bmm}	1.37 ± 0.33	0.56 ± 1.19	18.43
10	B_{bbb}	1.49 ± 0.03	-0.66 ± 0.08	13.32
10	B_{bbm}	1.48 ± 0.04	-0.69 ± 0.13	13.60
10	B_{bmm}	1.46 ± 0.09	-0.69 ± 0.29	14.13
6.7	B_{bbb}	1.36 ± 0.02	-0.74 ± 0.06	13.21
6.7	B_{bbm}	1.36 ± 0.03	-0.82 ± 0.09	12.87
6.7	B_{bmm}	1.32 ± 0.07	-0.86 ± 0.22	13.13

(13 degrees-of-freedom)

To avoid biasing our bias measurements,

We need to go beyond standard perturbation theory!

We adopt a new approach that addresses two issues:

1) Accurate modeling of non-linearities

2) Model dependence on smoothing

New approach to measuring actual halo(galaxy) bias

Pollack, Smith, Porciani 2012 (in preparation)

1) In place of SPT, use N-body simulations to measure full non-linear matter bispectrum and polyspectra up to sixth-order (or any order in expansion series) in the matter density

How? 'Bias-by-hand' The smoothed matter density method

2) The approach requires smoothing the matter density field in order to truncate series to second-order.

Trouble with smoothing: the local bias model assumes there is some filter scale where it is a valid approximation, but in actuality we do not know a priori what the smoothing scale is.

De-Smooth

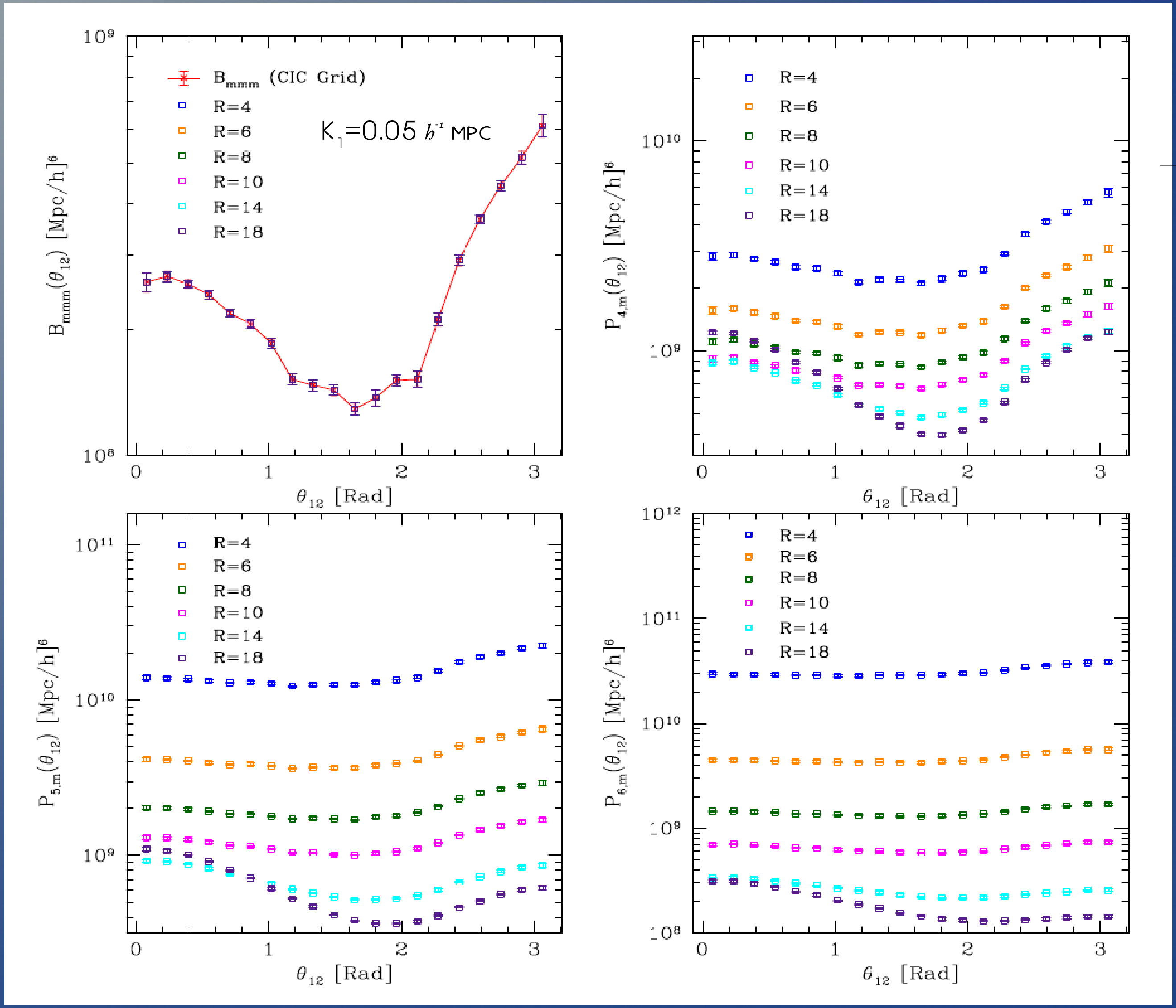
But, it doesn't fully undo smoothing

Bias parameter estimation for each R via maximum likelihood techniques performing a simultaneous fit of data with ensemble-averaged model

"Marginalize" over R to get $P(b_1, b_2)$

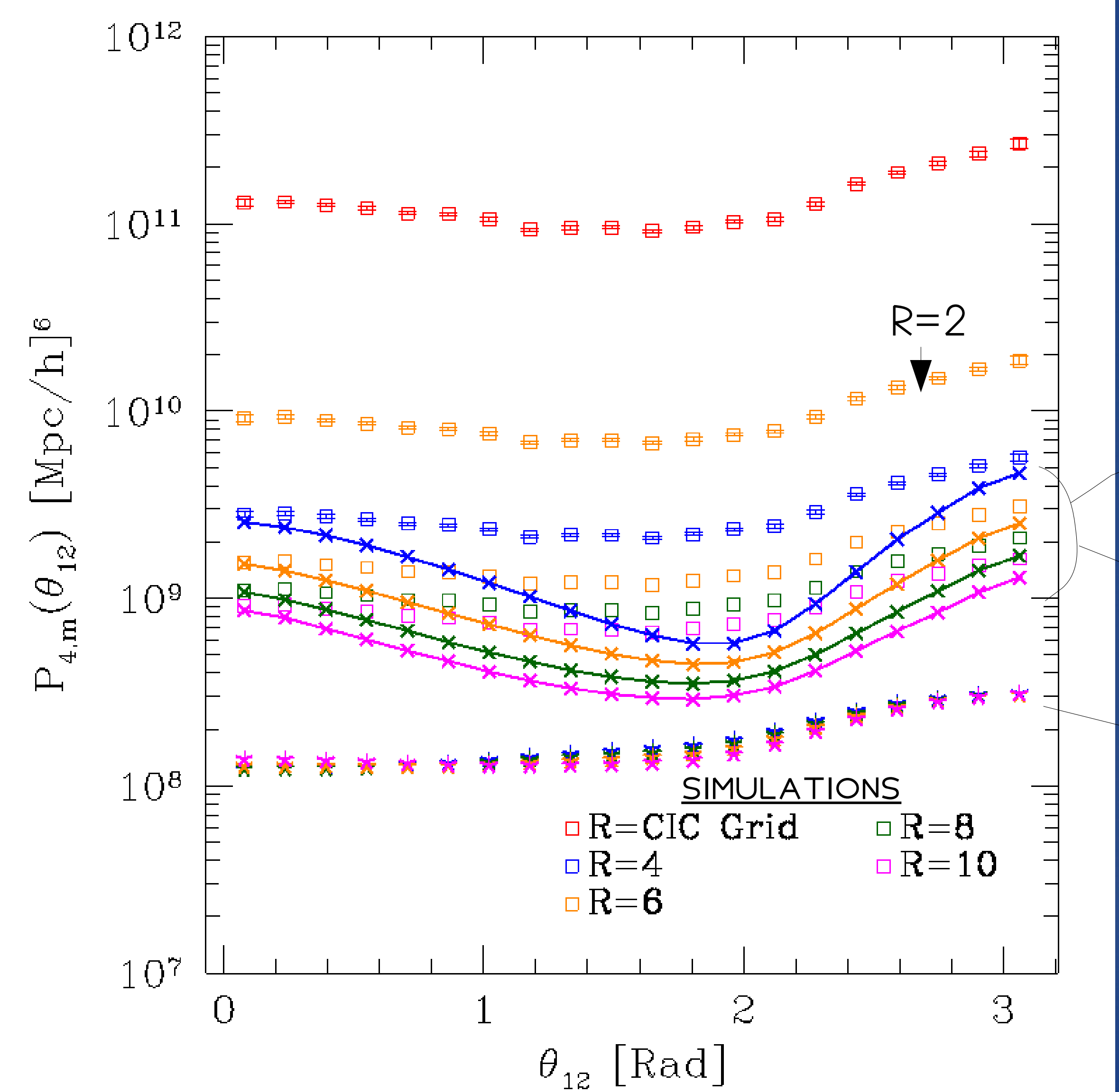
Consider range of R=2-18 Mpc/h in steps ΔR=0.01 Mpc/h

For n>3 P_N-terms, smoothing dependence remains even after applying de-smoothing operation



Diverging behavior as R→0

Understanding divergence



Diverging behavior as R→0

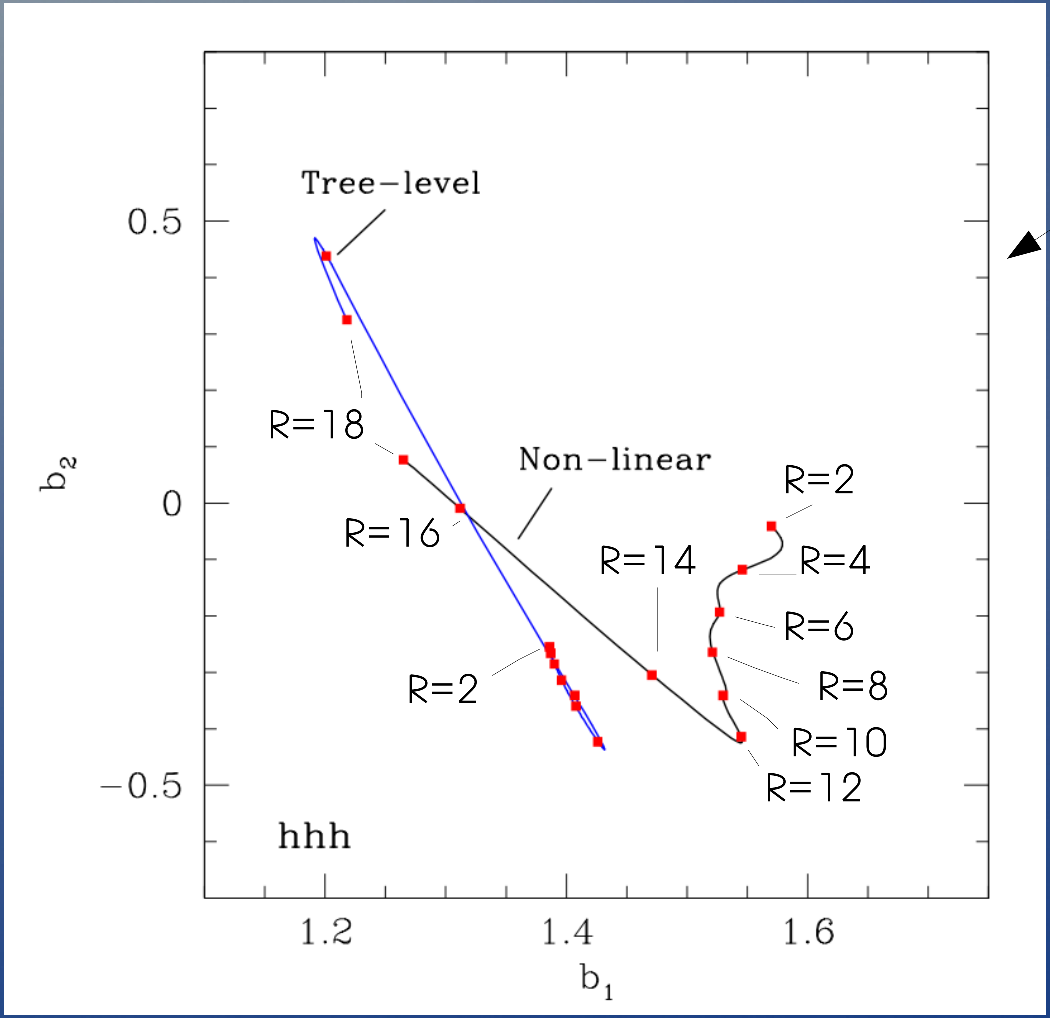
Attributed to connected part of poly-spectra

SPT
Integral over connected and unconnected tree-level trispectrum, $T(\boldsymbol{q}_1, \boldsymbol{k}_1 - \boldsymbol{q}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$

Sum of unconnected parts of trispectrum (i.e. cyclical products of the linear power spectrum)

Halo bias results comparisons
New approach vs. old 'tree-level' approach

Before Marginalization: "Running of the bias"



Maximum Likelihood Estimates of b_1 and b_2 for 1600 de-smoothed models each using tree-level and full non-linear terms

R	χ^2 (Tree)	χ^2(Non-Linear)
2	598.01	586.85
4	597.96	586.94
6	597.97	586.99
8	598.17	587.16
10	598.87	587.97
12	601.13	592.03
14	608.74	606.31
16	607.66	615.39
18	608.81	612.88

D.O.F.=598

χ^2 equally good, which means within this range there is no preferred model

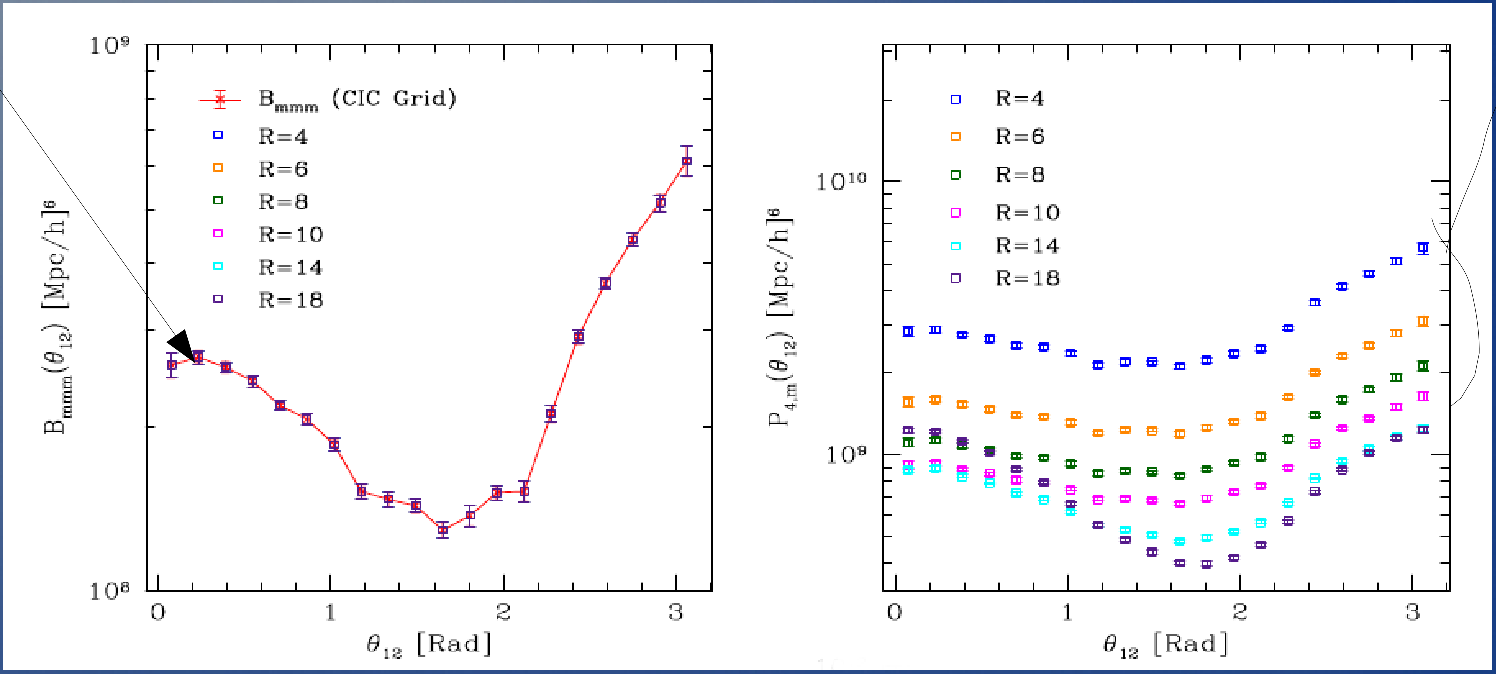
There's an explanation for this.

Consider fitting the halo-matter-matter bispectra with the following de-smoothed model:

$$\mathcal{B}_{\rm bmm} = b_1 \mathcal{B}_{\rm mmm} + \frac{b_2}{2}\mathcal{P}_{4,\rm m}$$

All de-smoothed B_{MMM} overlap

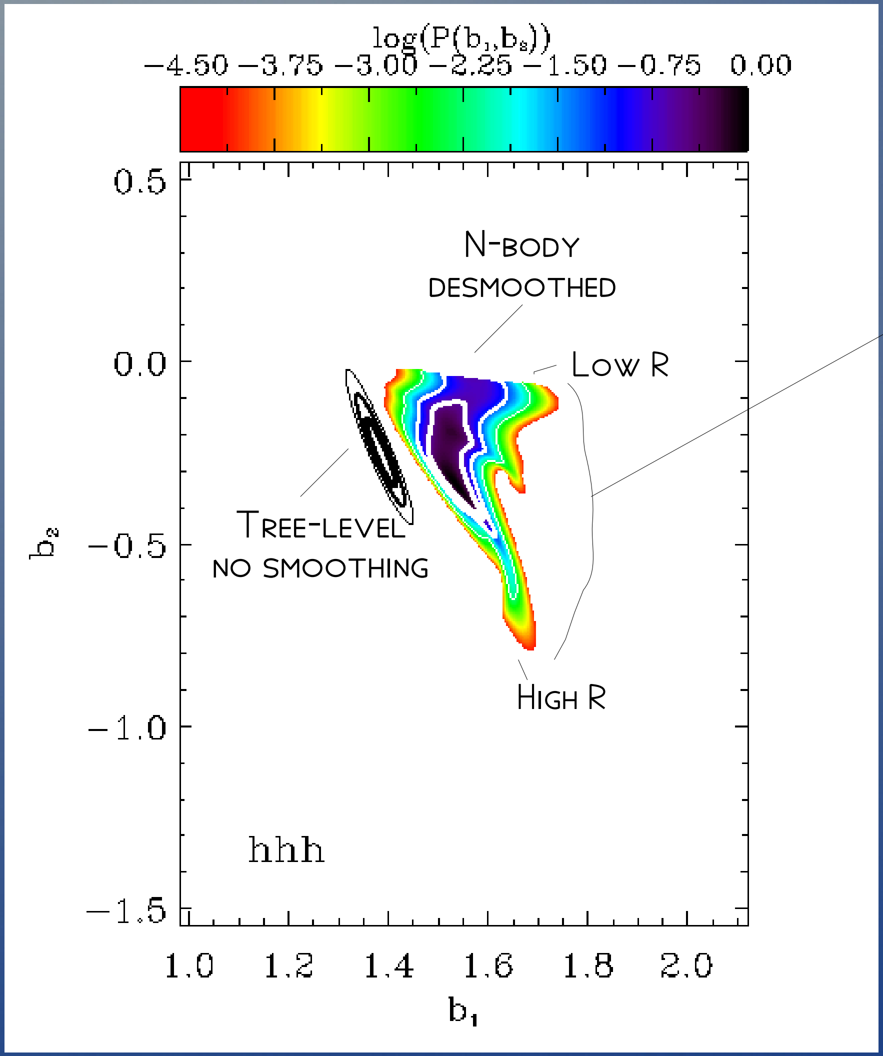
Ratios shows $P_{4,M}$ roughly scales as a constant



Confidence regions for R span slightly different parameter space as R is varied

There is a degeneracy between non-linear bias term, b_2 and filter scale R, which controls level of non-linearity in density field.

HALO BIAS RESULTS COMPARISONS
NEW APPROACH VS. OLD TREE-LEVEL APPROACH



Credibility initervals for 68.3%, 95.4%, and 99.73% are shown in lines of decreasing thickness

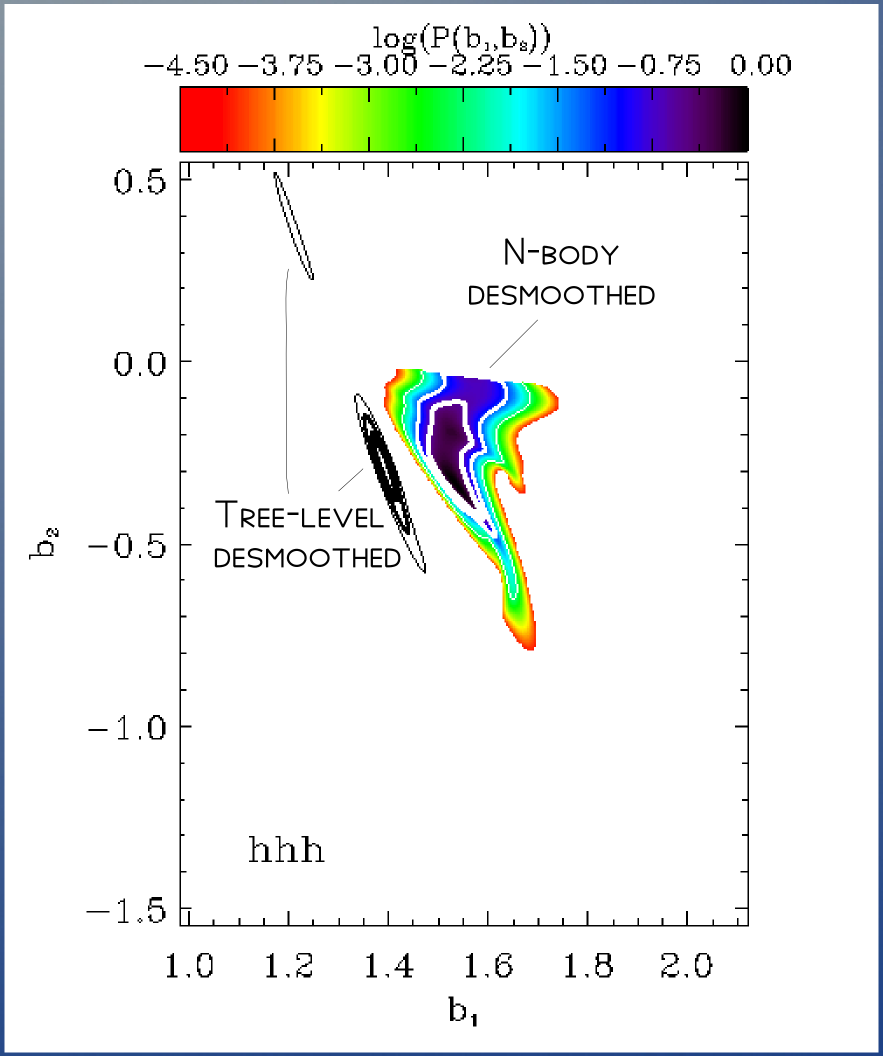
Multi-point solutions lead to wedge-shaped full marginal joint posterior PDF.

As R→0, b_1 becomes uncertain

As R→∞, b_2 becomes uncertain

Clear offset (or bias) between the results utilizing full non-linear model with de-smoothing dependence versus reliance on tree-level halo bias bispectrum model

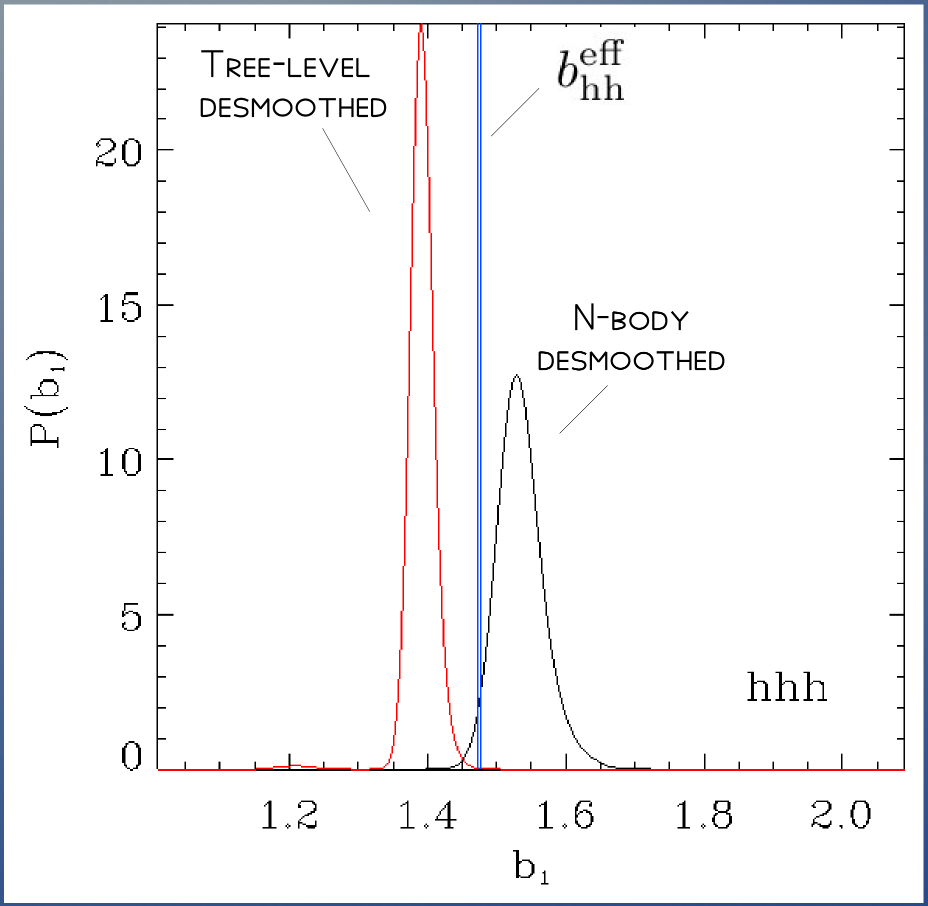
Halo Bias Results Comparisons
New Approach vs. Old Tree-Level Approach

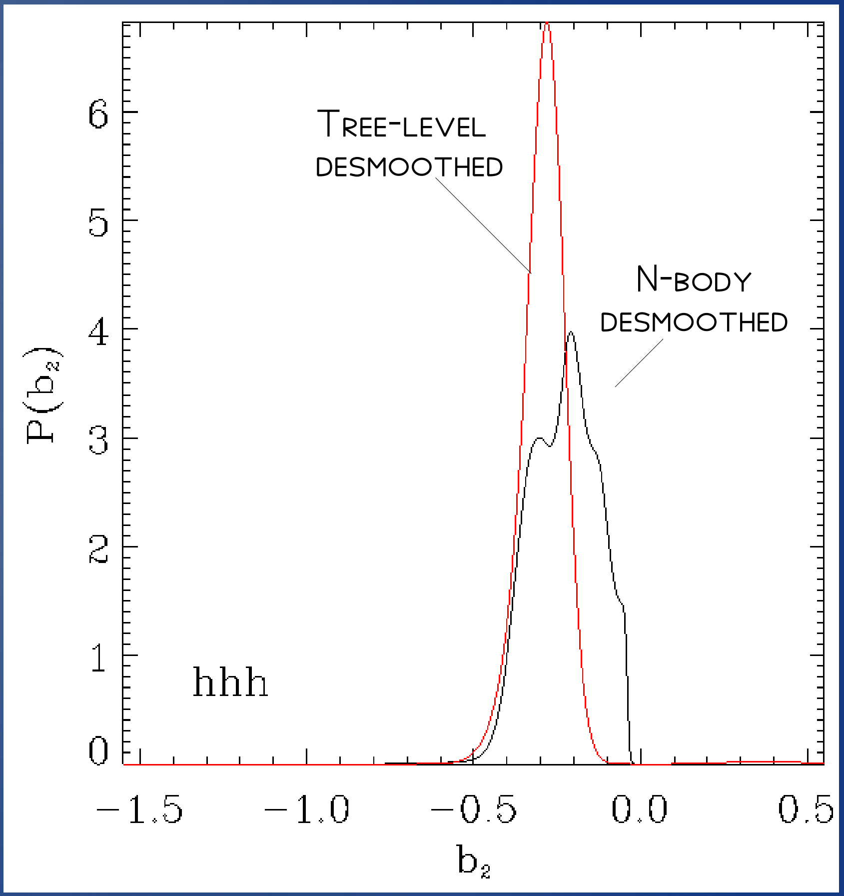


De-smoothed tree-level results suffers from oversmoothing.

3σ contour due to good fit results at large R>17 Mpc/h , so we ignore it.

Halo bias results comparisons
New approach vs. old tree-level approach





Nice constraint on b_1, but is it good enough?

There is weak overlap with b_{hh}^{eff}

Maybe, $b_{hh}^{eff} \neq b_1$?

Constraints on b_2 are weak.

Summary & Conclusions

- Non-linear bias determination based on standard perturbation theory gives a biased measurement of the actual halo (galaxy) bias.

- We developed a method based on our 'bias-by-hand' approach to measure non-linear poly-spectra accurately relying on the sole use of n-body simulations.

- Since models depend on smoothing, we relax this dependence by de-smoothing (i.e. dividing out the window functions) and "marginalize" our joint posterior probability over the filter scale R.

- The constraints for the joint PDF seem weak resulting from multi-point solutions that arise from a correlation between the filter scale R and the non-linear bias term, b_2.

- Marginal probabilities on the linear bias term, b_1 are favorable, while the constraints on b_2 are weak.

- If indeed $b_{eff}=b_1$, then the results may point towards inaccuracies in the local biasing scheme, which may need to account for non-local effects such as tidal fields, mergers and stochasticity. Other remedies include renormalizing the bias parameters.

- nevertheless, our implementation is a new & improved application of the local eulerian halo bias model that is free of the limitations of SPT and can be used to better test additional models of bias