On the Likelihood Analysis of the Large-Scale Halo Bispectrum

Andrea Oddo

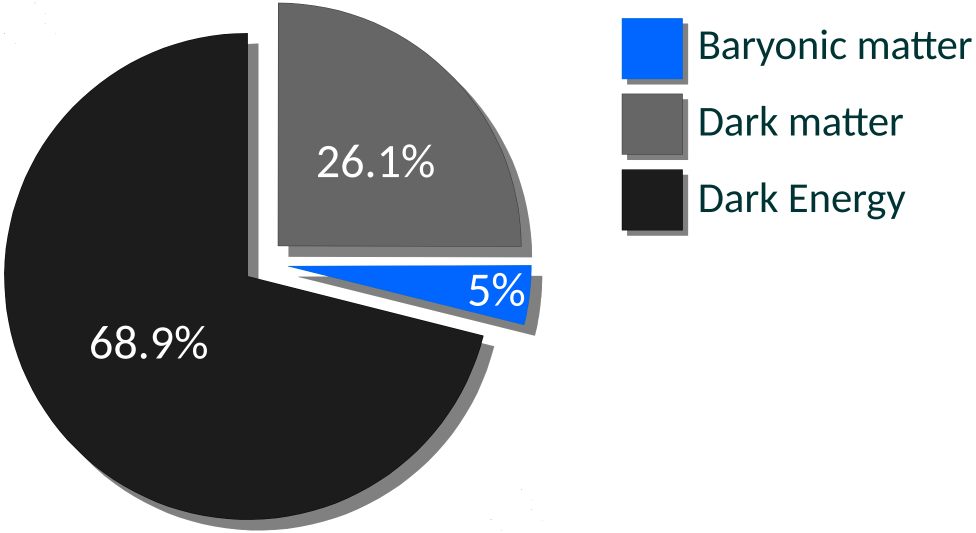
SISSA – Scuola Internazionale Superiore di Studi Avanzati
Astroparticle Physics Curriculum

In collaboration with:
Emiliano Sefusatti, Cristiano Porciani,
& Higher Order Statistics WP of the Euclid Galaxy Clustering SWG

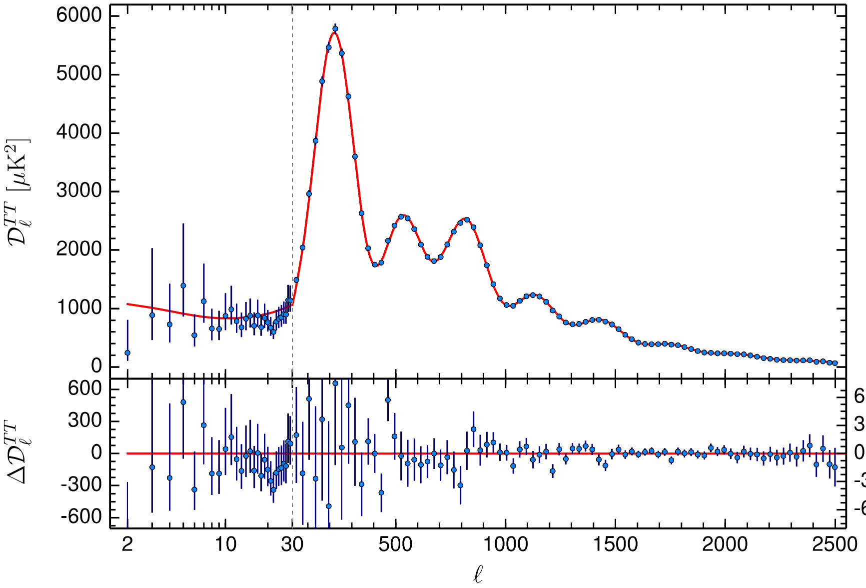
Astro@TS, @IFPU, Trieste
24th June 2019

The Standard ΛCDM Cosmology

Simplest model to reproduce observations with only **six parameters**.



95% of the Universe made of **unknown dark components**.



After Planck, we need information from 3D structures: LSS surveys (Euclid).

The Power Spectrum

Matter density contrast field

$$\delta_\mathrm{m}(\mathbf{x},t) = \frac{\rho_\mathrm{m}(\mathbf{x},t) - \bar{\rho}_\mathrm{m}(t)}{\bar{\rho}_\mathrm{m}(t)} \quad \overset{\mathcal{FT}}{\Rightarrow} \quad \delta_\mathbf{k}(t)$$

Power Spectrum

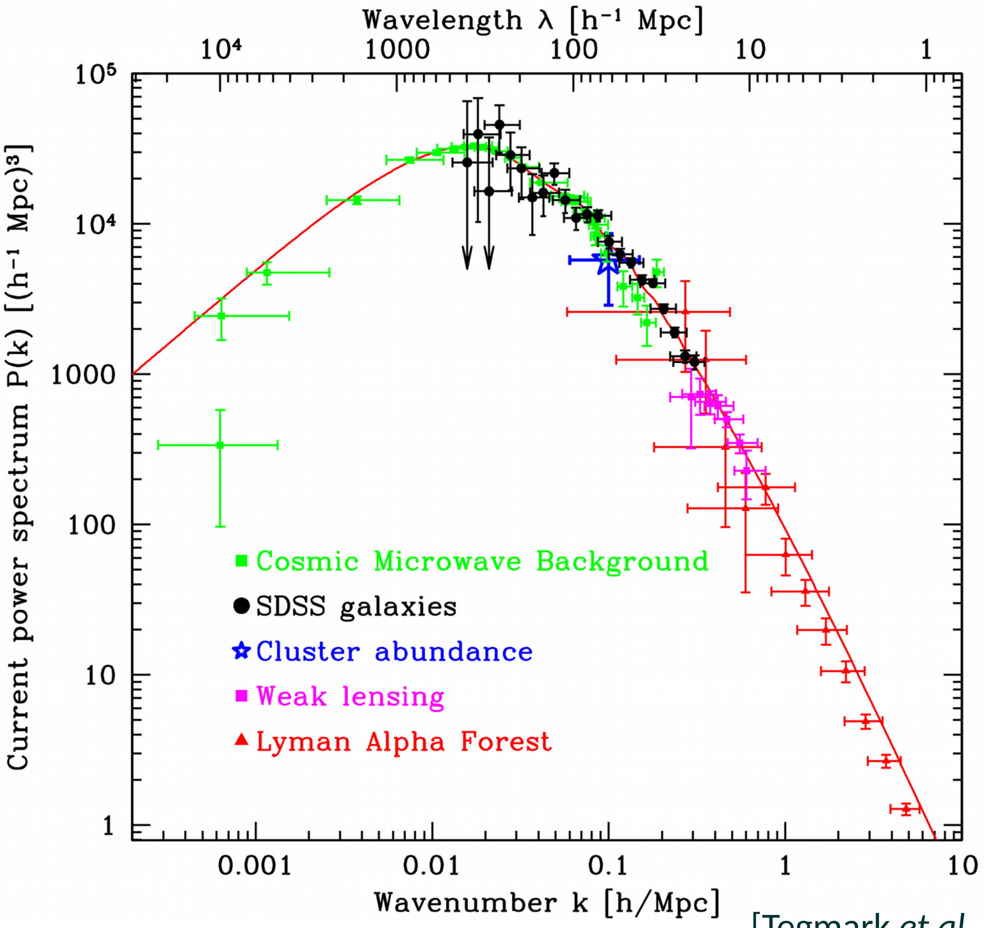
$$\delta_D(\mathbf{k}+\mathbf{k}')P_\mathrm{m}(k) = \langle \delta_\mathbf{k}\delta_{\mathbf{k}'} \rangle$$

It's related to the variance of the field.

What we actually observe is the galaxy density field:

$$\delta_\mathrm{g} = \delta_\mathrm{g}(\delta, \theta)$$

Galaxy power spectrum $P_\mathrm{g}(k)$



[Tegmark *et al.*, 2004]

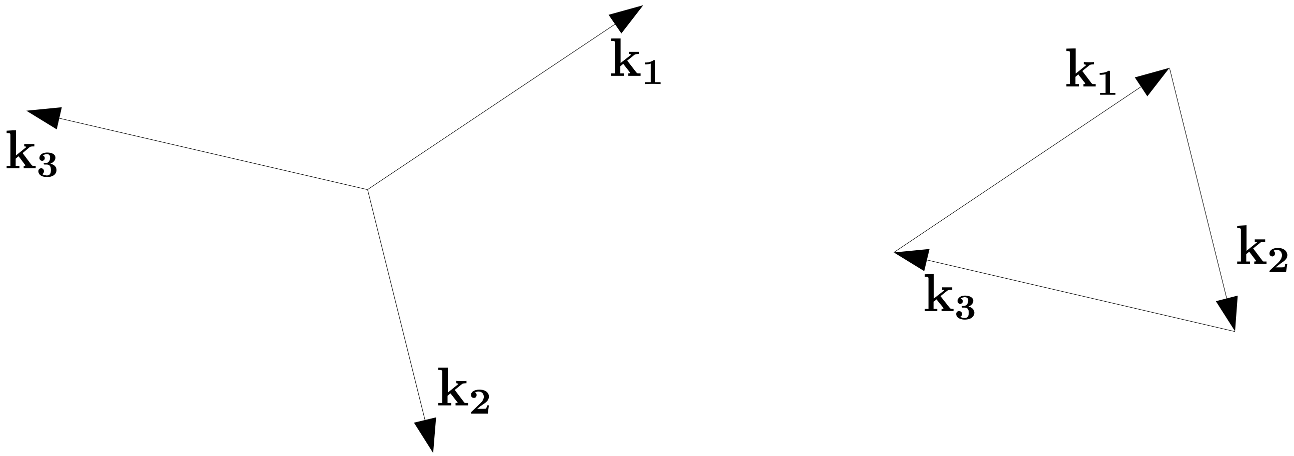
Non-gaussianities from non-linearities: Bispectrum

Mean $= 0$, variance $\sim P(k)$. What about higher-order moments?

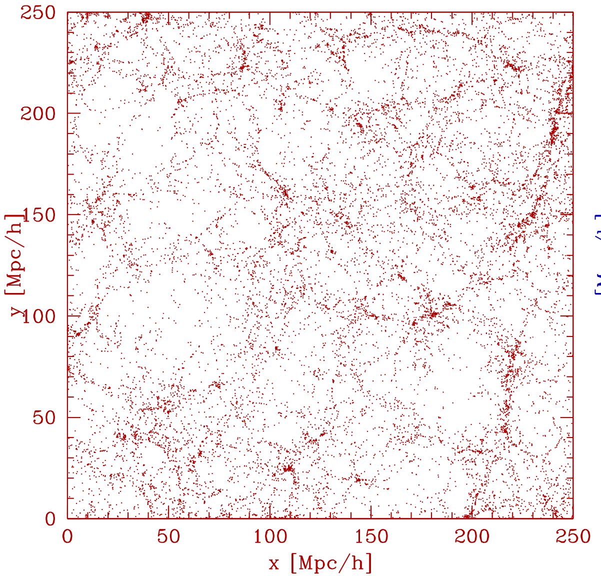
If field is gaussian: $\langle \delta^0_{\mathbf{k}_1} \dots \delta^0_{\mathbf{k}_N} \rangle = 0$ for $N > 2$

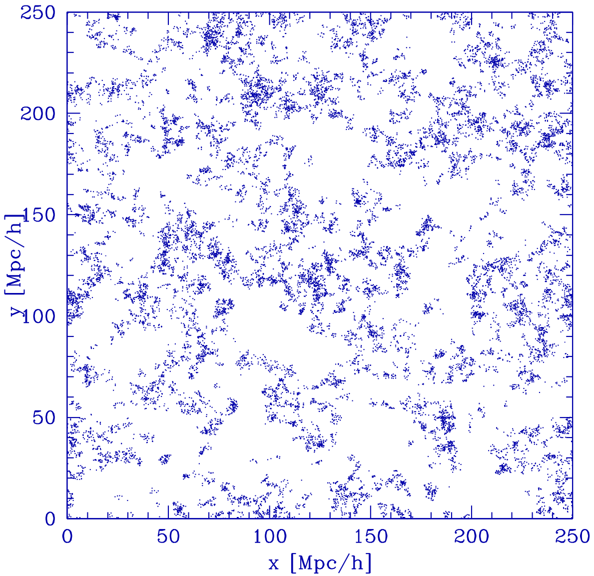
However, gravity is a non-linear process.
Non-linear evolution introduces non gaussianities: $\langle \delta^{NL}_{\mathbf{k}_1} \dots \delta^{NL}_{\mathbf{k}_N} \rangle \neq 0 \; \forall N$

We focus here on $N{=}3$: the **bispectrum**. $B(\mathbf{k_1}, \mathbf{k_2}, \mathbf{k_3}) \delta_D(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) \equiv \langle \delta_{\mathbf{k_1}} \delta_{\mathbf{k_2}} \delta_{\mathbf{k_3}} \rangle$

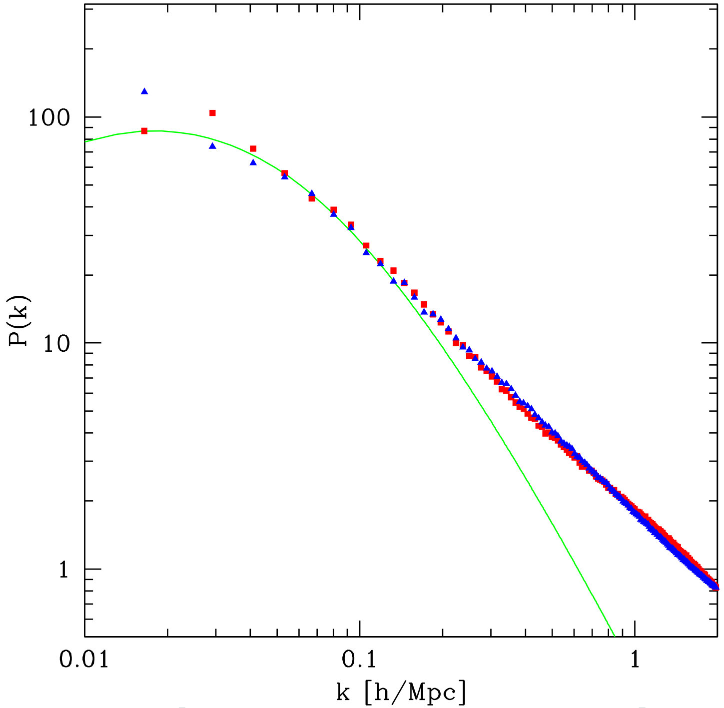


Motivation



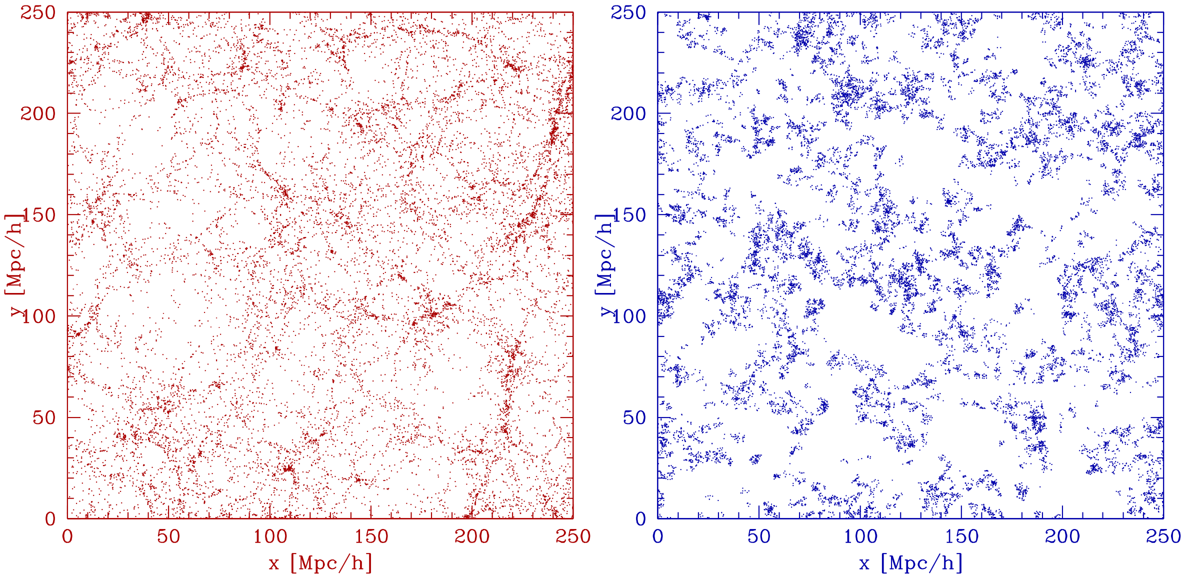


N-body vs **Levy flight**

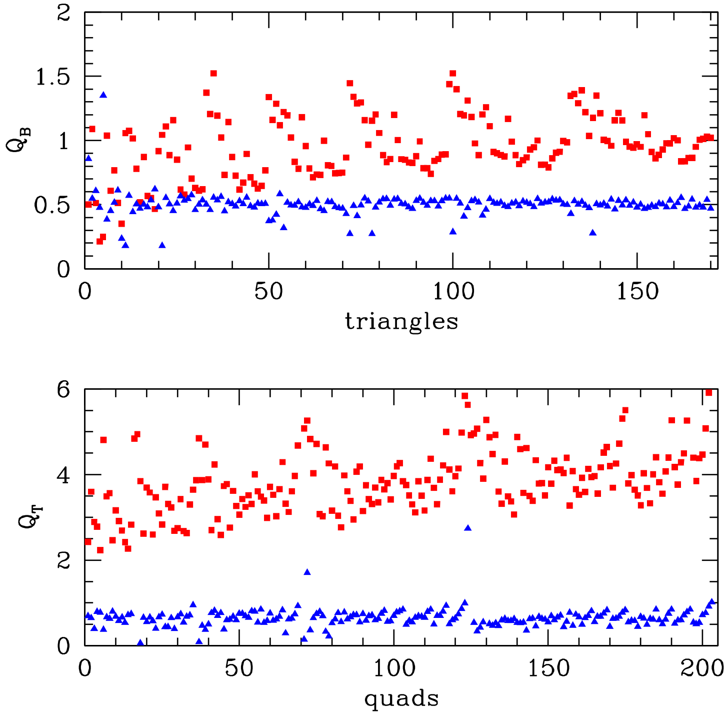


[Sefusatti & Scoccimarro, 2005]

Motivation

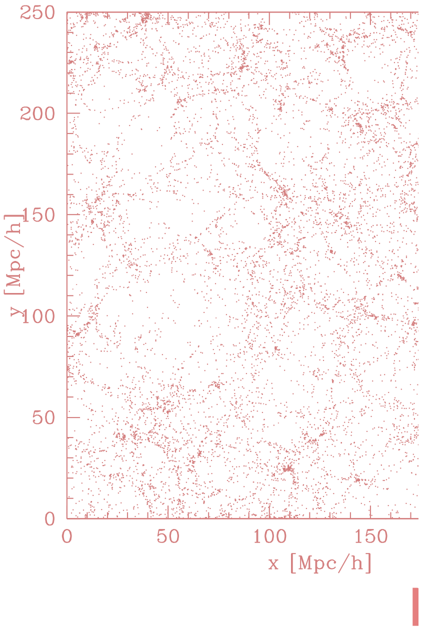


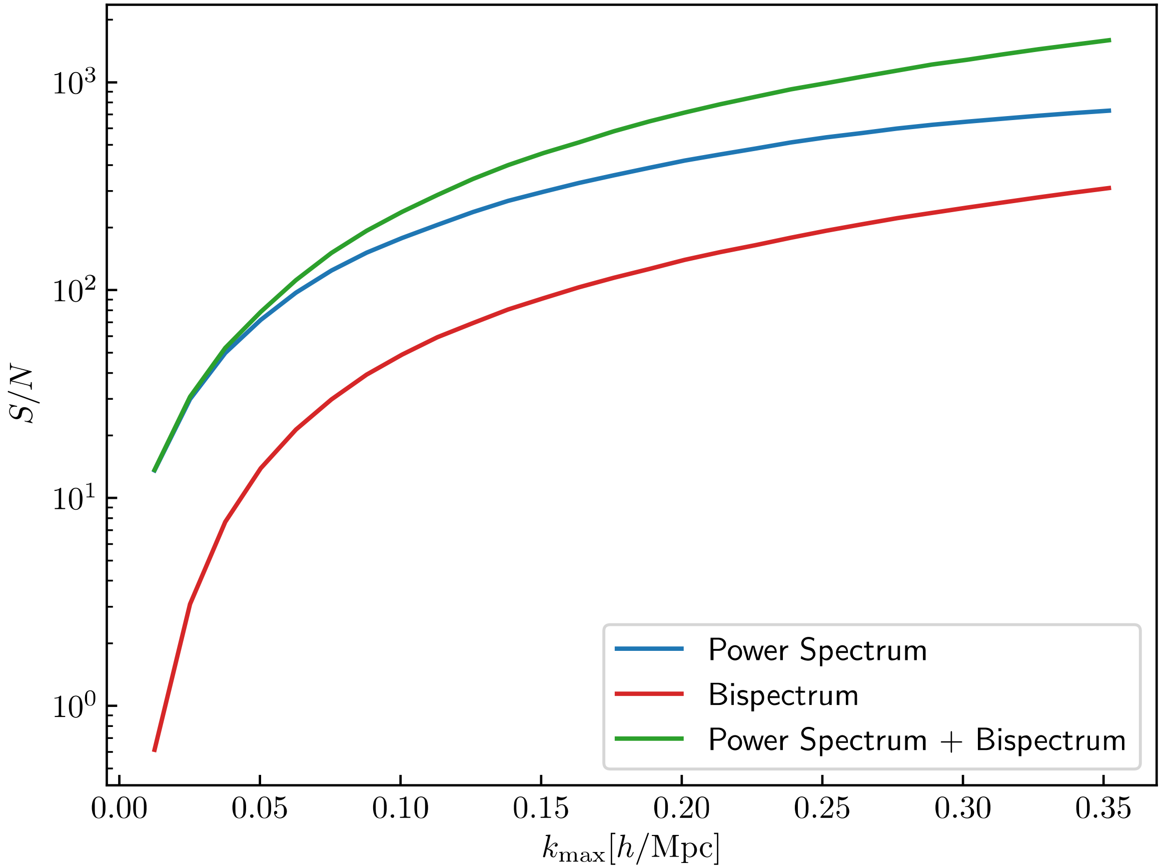
N-body vs **Levy flight**

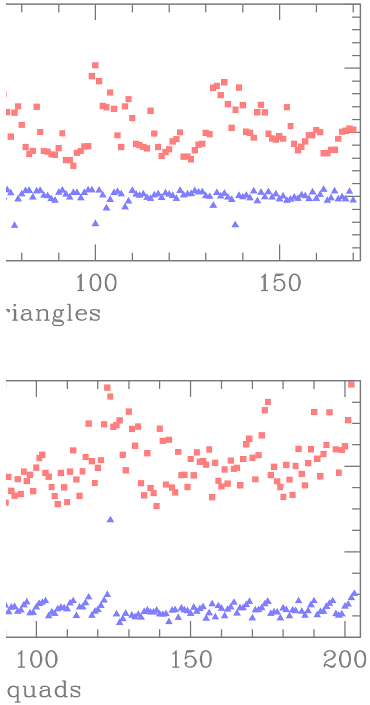


[Sefusatti & Scoccimarro, 2005]

Motivation







Scoccimarro, 2005]

Outline of the Project

Build a **likelihood** pipeline that can fit the **bispectrum**, accounting for different (usually neglected) effects that can introduce both statistical and/or systematic uncertainties.

Attention on proper definition of **goodness-of-fit** and on developing tools for **model selection**.

The likelihood has been tested on a large number (~300) of N-body simulations.

Results will appear in a soon-to-be-submitted paper [AO+2019]

Theoretical model

Fixed cosmological parameters; non-linear, non-local bias expansion for the halo field:

$$\delta_h = b_1\delta + \frac{1}{2}b_2\delta^2 + \gamma_2\mathcal{G}_2 + \gamma_3^-\Delta_3\mathcal{G}$$

δ is the full non-linear density field: $\delta_{\mathbf{k}_1} = \delta^{(1)}_{\mathbf{k}_1} + \delta^{(2)}_{\mathbf{k}_1} + \delta^{(3)}_{\mathbf{k}_1} + \dots$

Tree-level halo bispectrum:

$$B_h(\mathbf{k_1},\mathbf{k_2},\mathbf{k_3}) = b_1^3 B_{TL}(\mathbf{k_1},\mathbf{k_2},\mathbf{k_3}) + b_2 b_1^2\Sigma(\mathbf{k_1},\mathbf{k_2},\mathbf{k_3}) + 2\gamma_2 b_1^2 K(\mathbf{k_1},\mathbf{k_2},\mathbf{k_3}) +$$

$$+\left\langle\frac{1}{n}\right\rangle(1+\alpha_1)b_1^2\left[P_L(k_1)+P_L(k_2)+P_L(k_3)\right]+\left\langle\frac{1}{n^2}\right\rangle(1+\alpha_2)$$

Cosmology is only k-dependent terms, computed only once: **fast likelihood evaluation** in MCMC.

Five parameters: $\{b_1, b_2, \gamma_2, \alpha_1, \alpha_2\}$

Attention has to be paid on the way triangles are binned in the Fourier bins.

Covariance matrix

The covariance matrix has been estimated using a set of 10,000 mock simulations with the same cosmology of the N-body simulations.

We calibrated the halo catalogs in a way that the total power spectrum (including shot-noise) matched the one from the N-body.

$$C^B_{ij} \simeq \delta^K_{i_1 j_1}\, \delta^K_{i_2 j_2}\, \delta^K_{i_3 j_3}\, \frac{1}{N_B(t_i)} P_{tot}(k_{i_1})\, P_{tot}(k_{i_2})\, P_{tot}(k_{i_3}) + 5 \text{ perm.}$$

This gives the best match in the covariance matrix.

Large number of mocks is due to the large number of data-points (triangular Fourier bins) to fit.

Likelihood function

Multivariate gaussian likelihood:

$$\log \mathcal{L}_G = -\frac{1}{2}(\mathbf{B}_{\rm data} - \mathbf{B}_{\rm model})^T \cdot \mathbb{C}^{-1} \cdot (\mathbf{B}_{\rm data} - \mathbf{B}_{\rm model}) + {\rm const}$$

Anderson-Hartlap correction: $$\mathbb{C}^{-1} \rightarrow \frac{n-p-2}{n-1}\mathbb{C}^{-1}$$

To have a non-singular covariance matrix, $n > p$; bispectrum; $p = O(1000)$

The estimated covariance matrix is still a random object, sampled from a distribution

$$\log \mathcal{L}_{SH} = -\frac{n}{2}\log\left(1 + \frac{(\mathbf{B}_{\rm data} - \mathbf{B}_{\rm model})^T \cdot \mathbb{C}^{-1} \cdot (\mathbf{B}_{\rm data} - \mathbf{B}_{\rm model})}{n-1}\right) + {\rm const}$$

[Sellentin & Heavens (2016)]

$$\log \mathcal{L}_{\rm tot}(\{\mathbf{B}_\alpha\}|\boldsymbol{\theta}) = \sum_\alpha \log \mathcal{L}_\alpha(\mathbf{B}_\alpha|\boldsymbol{\theta})$$

Goodness of fit and model selection tools

Goodness of fit:

- **Chi-squared test** (frequentist concept)
- **Posterior predictive p-value** (ppp), fully Bayesian:

"Classical p-value averaged over the posterior distribution of parameters under the null hypothesis"

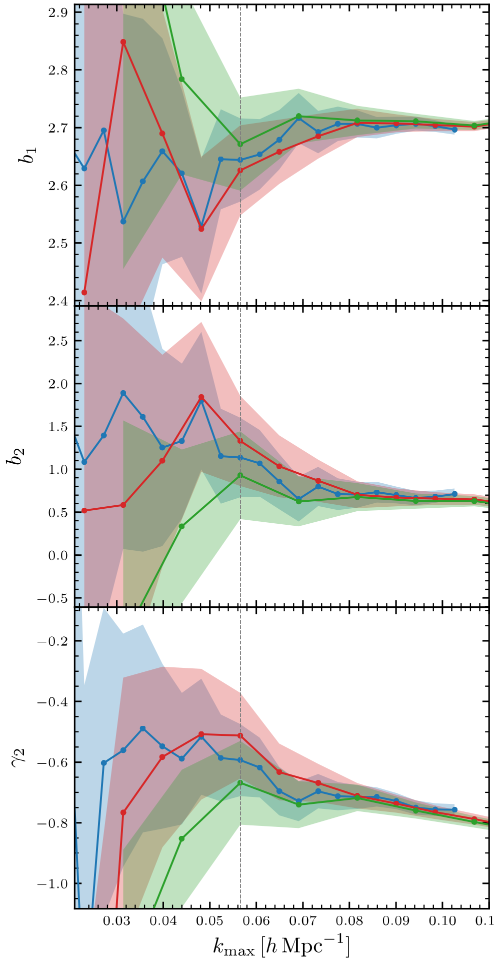
Model selection:

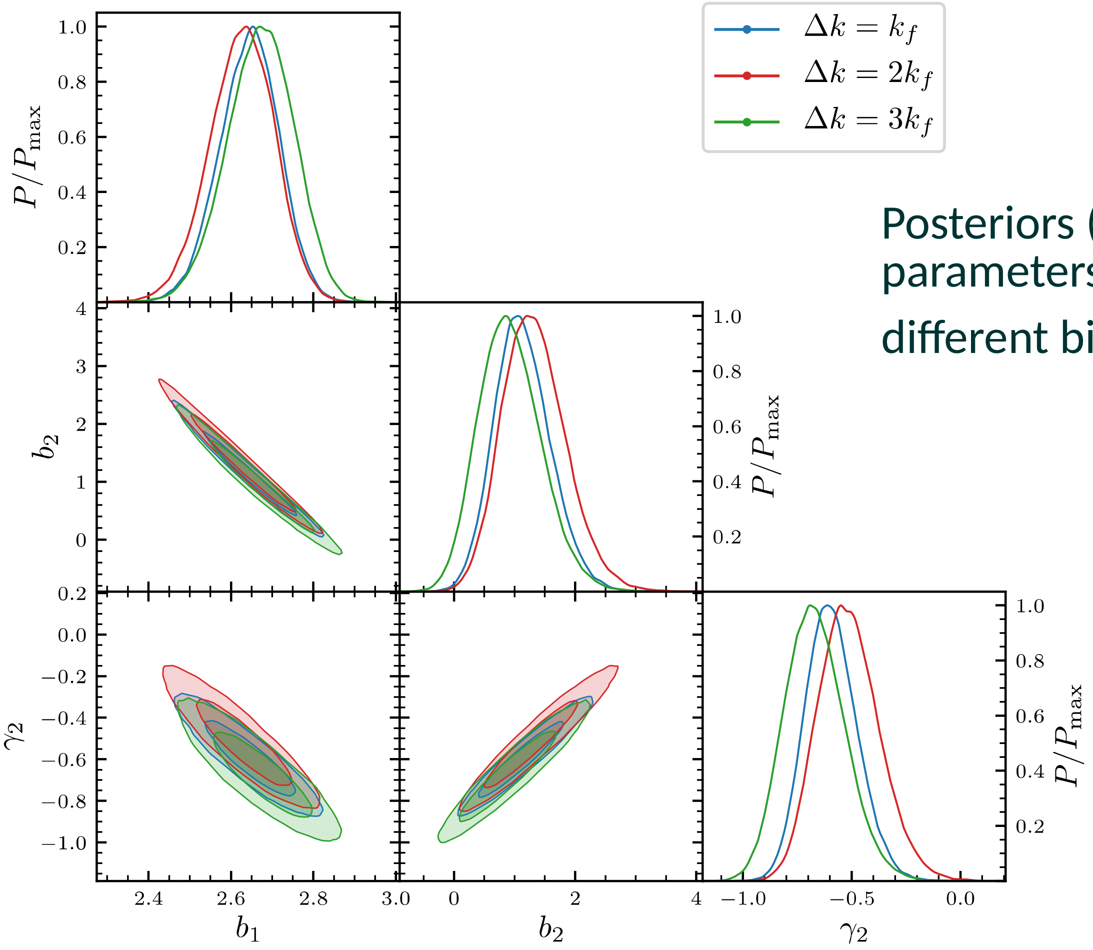
- **Deviance Information Criterion** (DIC)

$$\mathrm{DIC} = \frac{1}{2}\mathrm{var}(\mathrm{D}(\theta)) + \langle D(\theta) \rangle, \quad D(\theta) = -2\log\mathcal{L}(\theta)$$

- **Savage-Dickey density ratio** to compute the Bayes Factor

Results: Benchmark analysis, parameters

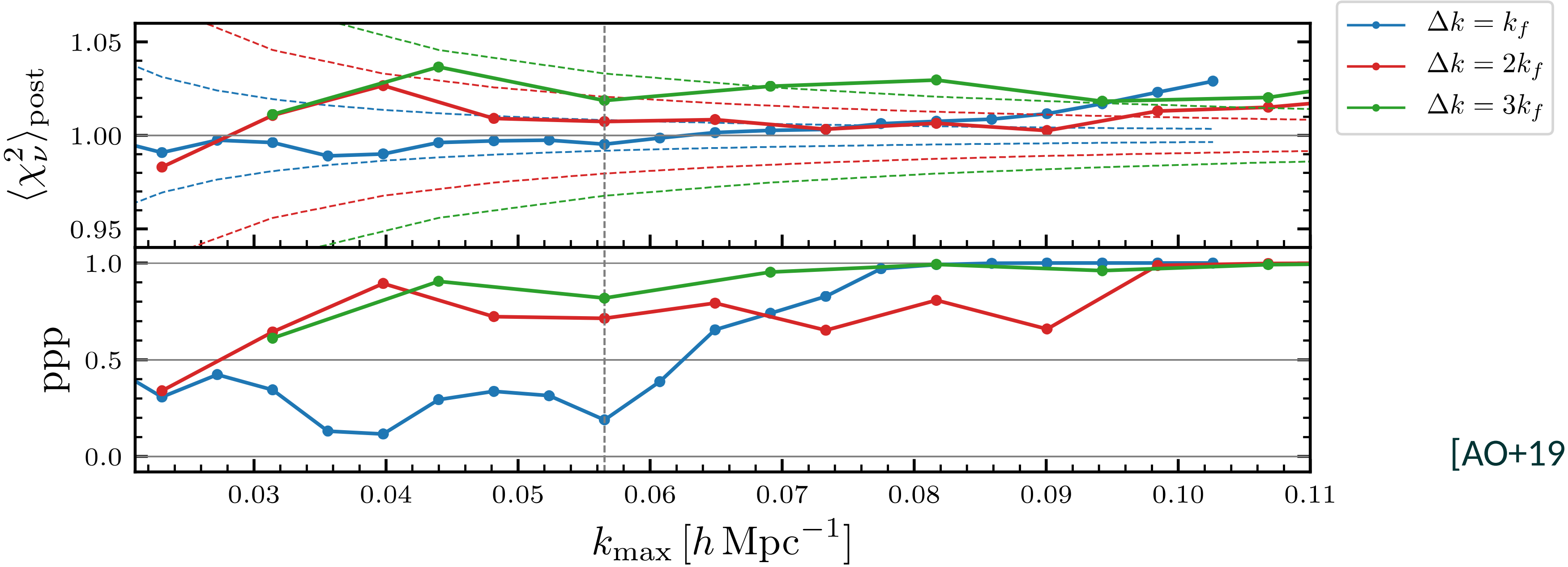




Posteriors (1D and 2D) for the parameters are consistent for different bin widths, at all $k_{\max}$.

[AO+19 in prep.]

Results: Benchmark analysis, goodness of fit



[AO+19 in prep.]

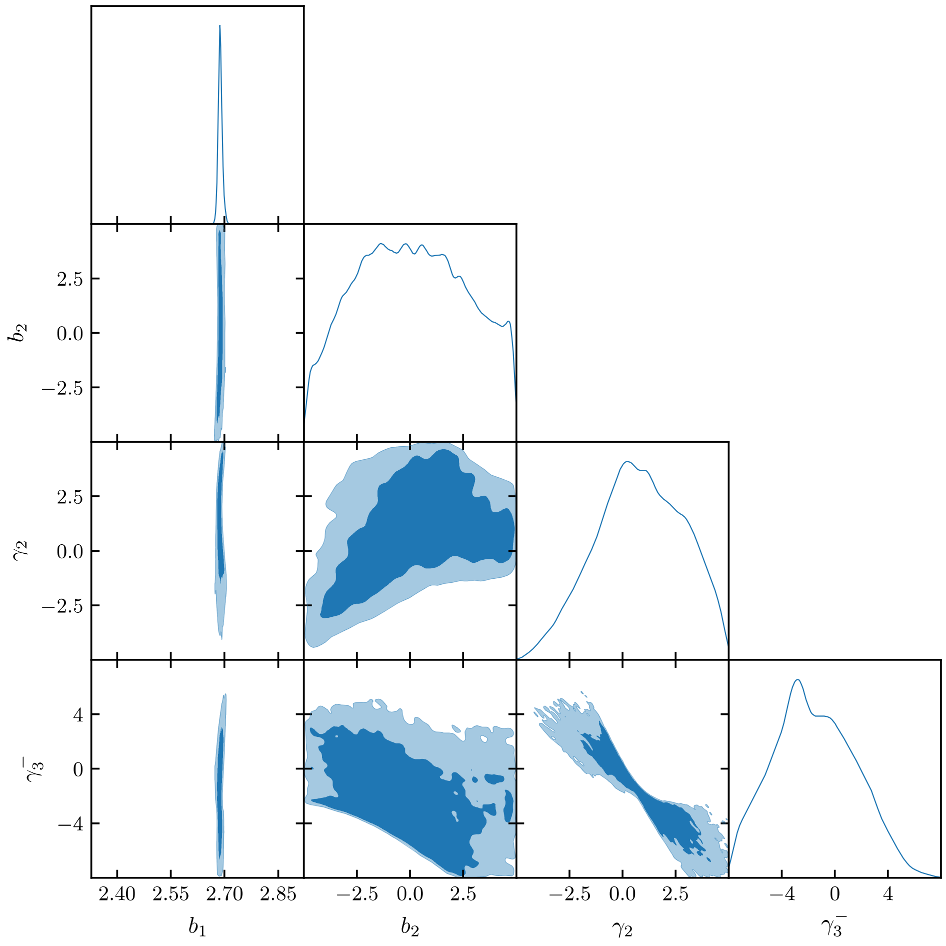
The $\langle\chi^2_\nu\rangle_{\rm post}$ and the $\rm ppp$ give similar results. The model appears to be consistent with the data up to $\sim 0.075h\ {\rm Mpc}^{-1}$, but the range of validity extends to $\sim 0.095h\ {\rm Mpc}^{-1}$ for $\Delta k = 2k_{\rm f}$.

Power spectrum + bispectrum: what would we gain?

$k_{\mathrm{max}} \sim 0.06\ h\ \mathrm{Mpc}^{-1}$

[AO+??, preliminary]

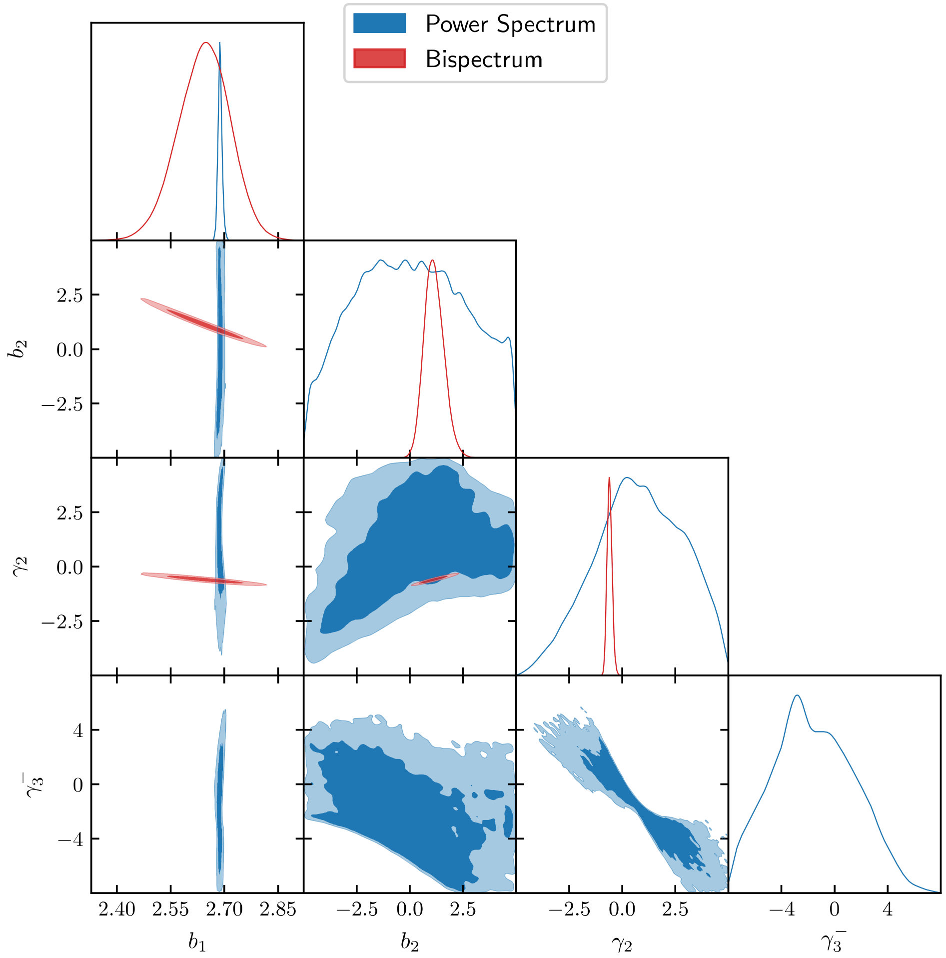
Power spectrum + bispectrum: what would we gain?



$k_{\rm max} \sim 0.06\ h\ \mathrm{Mpc}^{-1}$

[AO+??, preliminary]

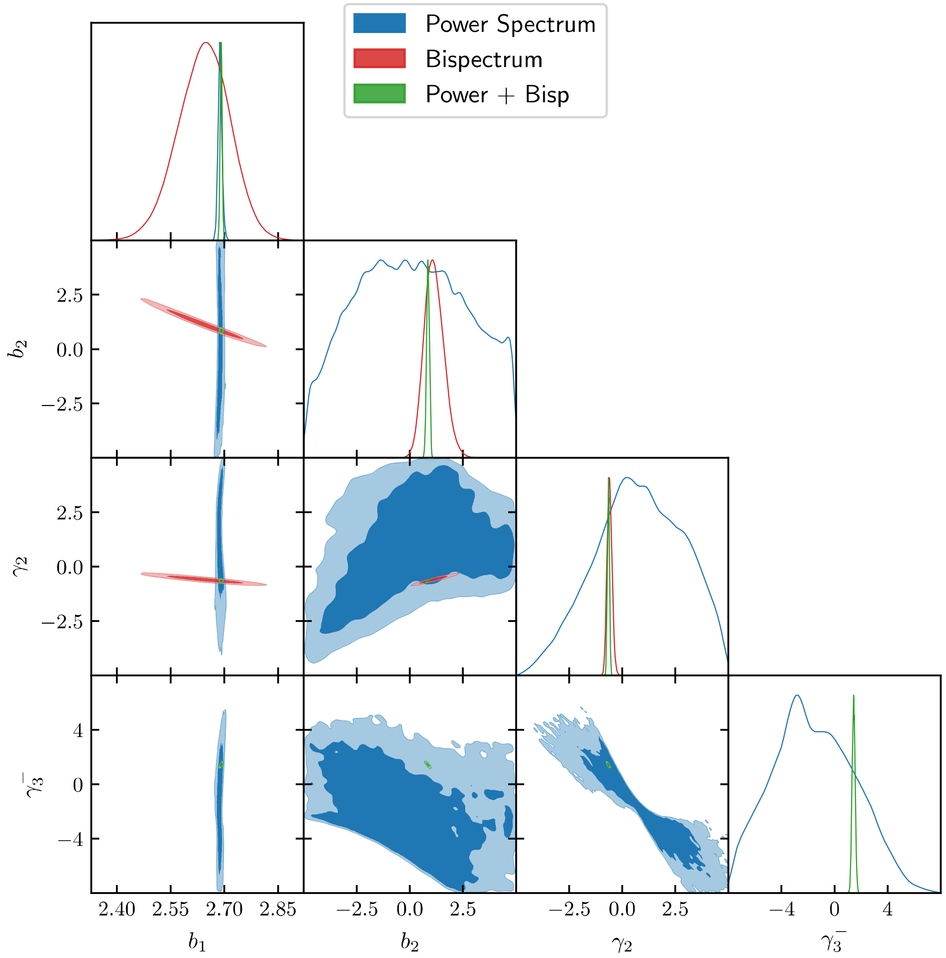
Power spectrum + bispectrum: what would we gain?



$k_{\rm max} \sim 0.06\ h\ {\rm Mpc}^{-1}$

[AO+??, preliminary]

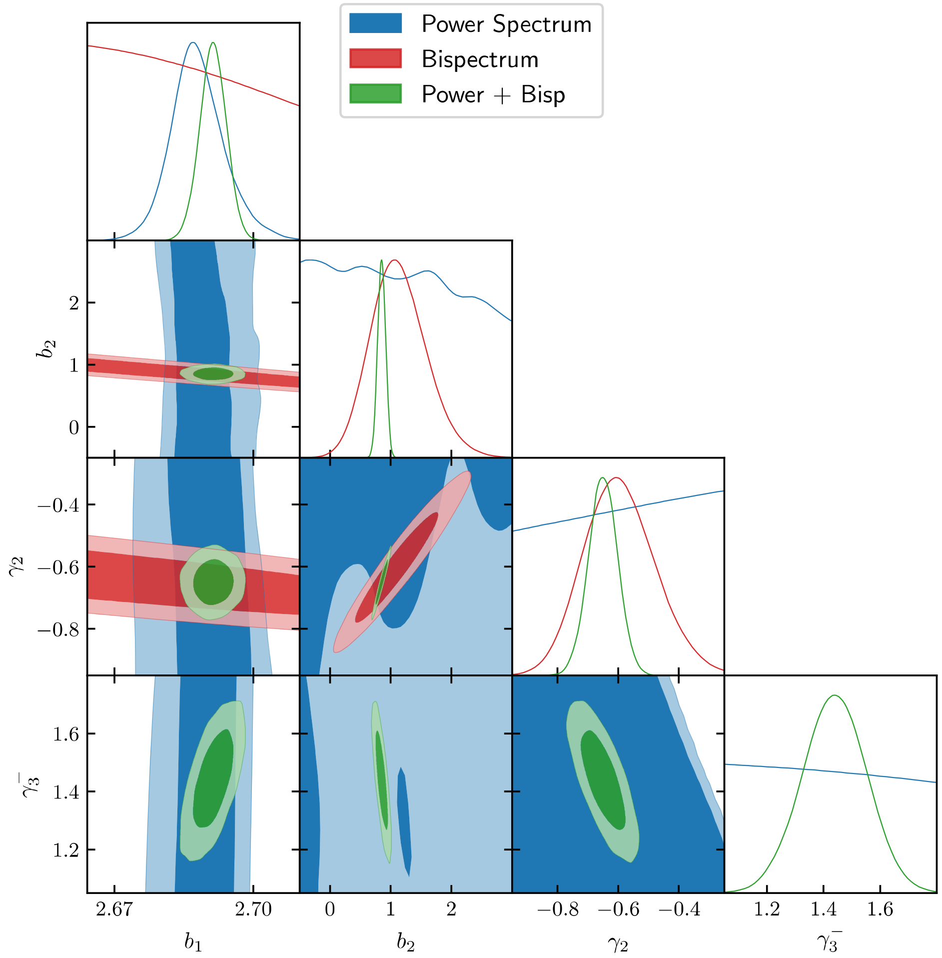
Power spectrum + bispectrum: what would we gain?



$k_{\rm max} \sim 0.06\ h\ \mathrm{Mpc}^{-1}$

[AO+??, preliminary]

Power spectrum + bispectrum: what would we gain?



$k_{\mathrm{max}} \sim 0.06\ h\ \mathrm{Mpc}^{-1}$

The tree-level bispectrum alone is not able to put tight constraints on b_1.
The 1-loop power spectrum alone is not able to constrain at all the non-linear e non-local bias parameters at large scales.
Using both, we can reduce the costraints an all bias parameters!

[AO+??, preliminary]

Summary & Conclusions

Galaxy bispectrum contains information not present in power spectrum and that we can extract.

We have built a likelihood for the robust inference of parameters from the bispectrum, accounting for effects introducing both statistical and systematic uncertainties.

Goodness of fit shows a good agreement between theory and data up to $\sim 0.075 h\ \mathrm{Mpc}^{-1}$

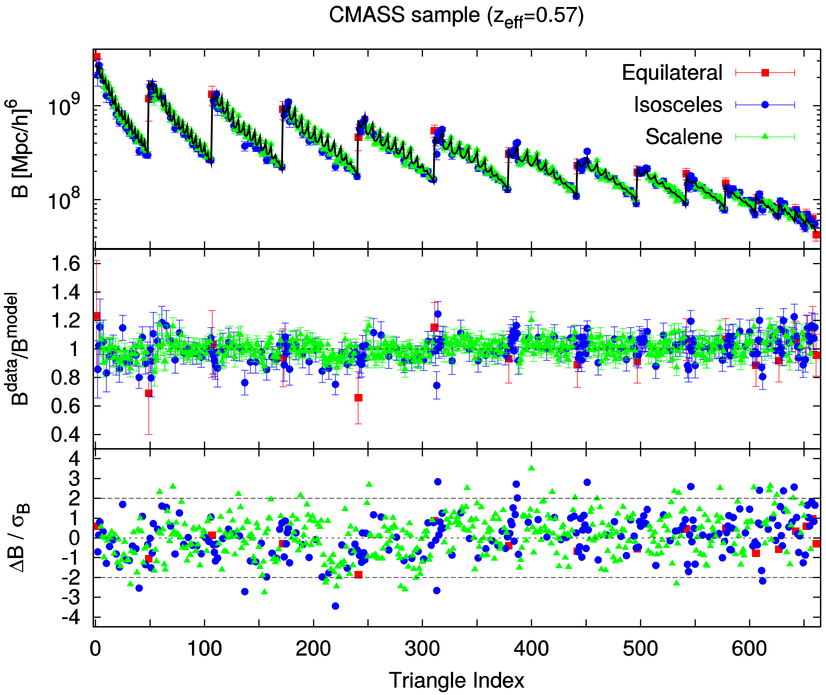
Model selection can help us in reducing the dimensionality of the parameter space.

Bispectrum can help in constraining parameters even at large scales.

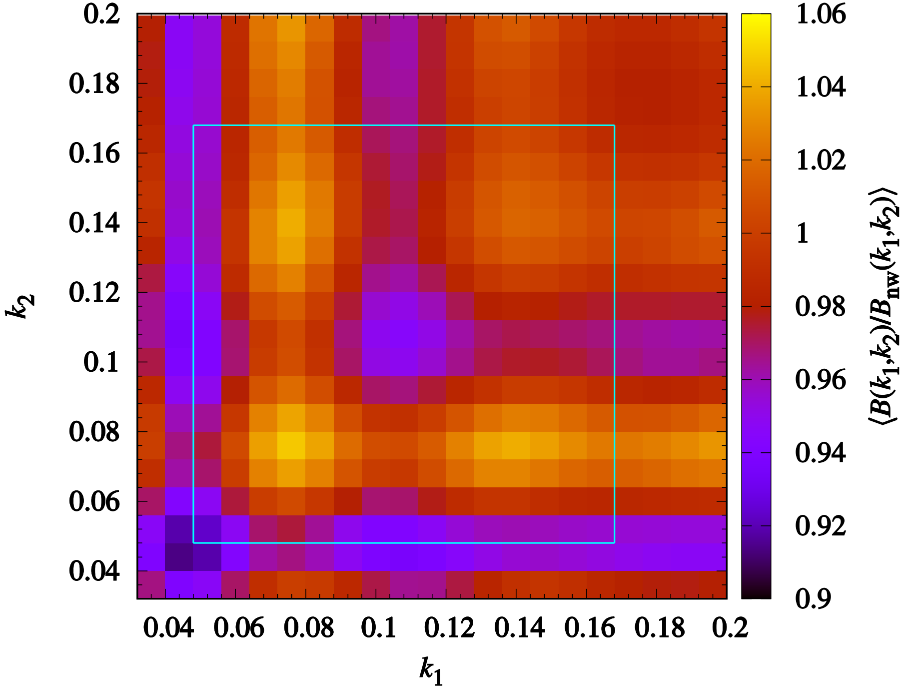
Thank you for your attention!

State of the Art

Gil-Marin *et al.* (2016):
Bispectrum in BOSS analysis



Pearson & Samushia (2017):
BAO detection in the bispectrum



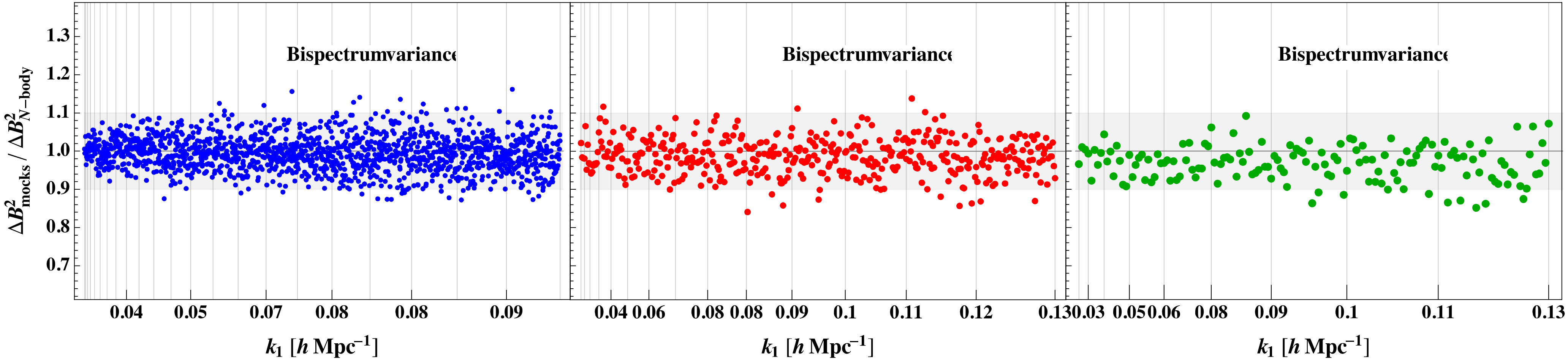
However, the models used are not valid at the level of future precision surveys, such as Euclid.
Moreover, covariance not properly accounted for (if any at all).

Covariance

The gaussian component of the bispectrum variance is

$$C^B_{ij} \simeq \delta^K_{i_1 j_1}\, \delta^K_{i_2 j_2}\, \delta^K_{i_3 j_3}\, \frac{1}{N_B(t_i)} P_{tot}(k_{i_1})\, P_{tot}(k_{i_2})\, P_{tot}(k_{i_3}) + 5 \text{ perm.}$$

It is important to match the ***total*** power spectrum (including shot-noise) of mocks and N-body.
We choose the minimum halo mass of the mocks in such a way that this happens.



[AO+19 in prep.]

The cross-correlations are consistent with the ones from the N-body, up to a lower noise.

Binning effects

Measurements are performed over bins.

$$(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \in (k_1, k_2, k_3) \text{ if } |\mathbf{q}_i| \in \left[k_i - \frac{\Delta k}{2}; k_i + \frac{\Delta k}{2}\right) \forall i = 1, 2, 3$$

Model must be properly **averaged** over the bin as well.

$$B_{\rm th}(k_1, k_2, k_3) = \frac{1}{N_T(k_1, k_2, k_3)} \sum_{\mathbf{q}_1 \in k_1} \sum_{\mathbf{q}_2 \in k_2} \sum_{\mathbf{q}_3 \in k_3} \delta_K(\mathbf{q}_{123}) B_h(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$$

Computationally expensive! Almost impractical when cosmology is changed. Alternative:

$$k_{{\rm eff},j} = \frac{1}{N_T(k_1, k_2, k_3)} \sum_{\mathbf{q}_1 \in k_1} \sum_{\mathbf{q}_2 \in k_2} \sum_{\mathbf{q}_3 \in k_3} q_j \delta_K(\mathbf{q}_{123}), \quad B_{\rm th}(k_1, k_2, k_3) = B_h(k_{{\rm eff},1}, k_{{\rm eff},2}, k_{{\rm eff},3})$$

However, this can differ from the properly averaged model by ~20%.

Roughly speaking, similar to $\langle f(x) \rangle$ vs. $f(\langle x \rangle)$.

Also possible to evaluate the model at the central triangle (k_1, k_2, k_3): large deviations!

Model selection: Deviance Information Criterion

We define the deviance as:

$$D(\text{data}, \boldsymbol{\theta}_j, M_j) = -2\log\mathcal{L}(\text{data}|\boldsymbol{\theta}_j, M_j) + 2\log C(\text{data})$$

The Deviance Information Criterion DIC is then

$$\text{DIC} = \langle D(\boldsymbol{\theta})\rangle_{\text{post}} + \frac{1}{2}\text{Var}(D(\boldsymbol{\theta}))$$

The second term is an estimator of the effective number of parameters, and represent a penalization for models with a large number of parameters.

We are interested in the difference of DIC between two models, ΔDIC.

In general, the model with the higher DIC is disfavoured, with $|\Delta\text{DIC}| \geq 5$ being substantial.

Model selection: Savage-Dickey density ratio

In order to evaluate the evidence that the data provide in favour of a model M_j w.r.t model M_k, we define the Bayes factor:

$$\mathrm{BF}_{jk} = \frac{p(\mathrm{data}|M_j)}{p(\mathrm{data}|M_k)} = \frac{\int \mathcal{L}(\mathrm{data}|\boldsymbol{\theta}_j)\,\pi(\boldsymbol{\theta}_j|M_j)\,\mathrm{d}\boldsymbol{\theta}_j}{\int \mathcal{L}(\mathrm{data}|\boldsymbol{\theta}_k)\,\pi(\boldsymbol{\theta}_k|M_k)\,\mathrm{d}\boldsymbol{\theta}_k}$$

In case of **properly nested models**, the Bayes factor can be computed using the Savage-Dickey density ratio

$$\mathrm{BF}_{jk} = \frac{\int P(\boldsymbol{\theta}_j, \boldsymbol{\psi} = \mathbf{c}|M_k)\,\mathrm{d}\boldsymbol{\theta}_j}{\int \pi(\boldsymbol{\theta}_j, \boldsymbol{\psi} = \mathbf{c}, |M_k)\,\mathrm{d}\boldsymbol{\theta}_j}$$

$\mathrm{BF}_{jk} > 3,\ 10,\ 100$ ⇒ substantial, strong, and decisive evidence against M_k (“Jeffreys scale”).

Binning effects

Measurements are performed over bins; model must be properly **averaged** over the bin as well.

$$B_{\rm th}(k_1, k_2, k_3) = \langle B_h(\triangle)\rangle \qquad \{\triangle \in (k_1, k_2, k_3)\}$$

Computationally expensive! Almost impractical when cosmology is changed. Alternative:

$$B_{\rm th}(k_1, k_2, k_3) = B_h\left(\langle\triangle\rangle\right) \qquad \{\triangle \in (k_1, k_2, k_3)\}$$

However, this can differ from the properly averaged model by ~20%.

Roughly speaking, similar to $\langle f(x)\rangle$ vs. $f(\langle x\rangle)$.

Also possible to evaluate the model at the central triangle (k_1, k_2, k_3): large deviations!

Goodness-of-fit: chi-squared

More of a frequentist approach, used as a reference:

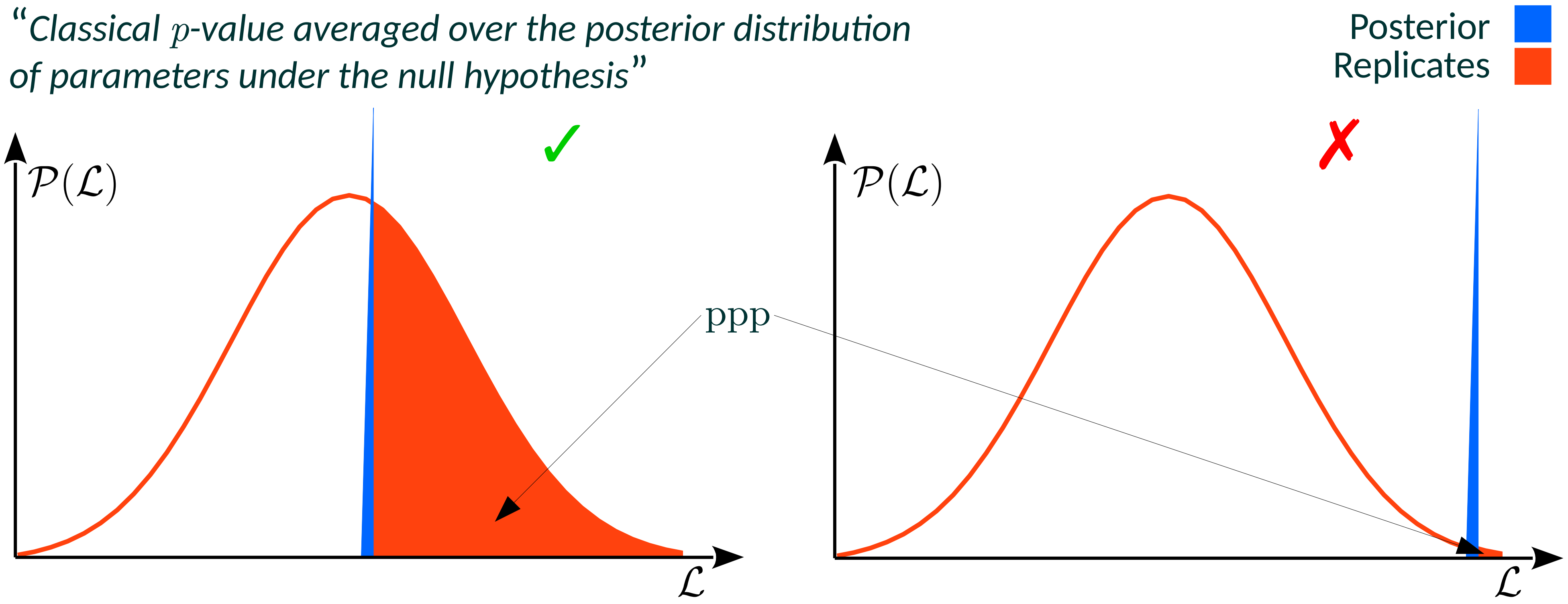
$$\langle \chi^2_\nu \rangle_{\text{post}} = \frac{1}{N_R N_T} \left\langle \sum_{r=1}^{N_R} (\mathbf{B}_r^{\text{data}} - \mathbf{B}^{\text{model}})^T \cdot \mathbb{C}^{-1} \cdot (\mathbf{B}_r^{\text{data}} - \mathbf{B}^{\text{model}}) \right\rangle_{\text{post}}$$

A fit as “good” if this chi-squared is between the values corresponding to p-values of 0.05 and 0.95, provided the number of degrees of freedom $N_R N_T$.

Evaluated as a function of k_{max}, it gives the range of scales in which the theory is valid.

Goodness-of-fit: Posterior Predictive *p*-value

"*Classical p-value averaged over the posterior distribution of parameters under the null hypothesis*"



The fit is "bad" if ppp is too close to either 0 or 1.
Evaluated as a function of k_{max}, it gives the range of scales in which the theory is valid.

Model selection tools

Deviance: $D(\mathrm{data}, \boldsymbol{\theta}_j, M_j) = -2\log\mathcal{L}(\mathrm{data}|\boldsymbol{\theta}_j, M_j) + 2\log C(\mathrm{data})$

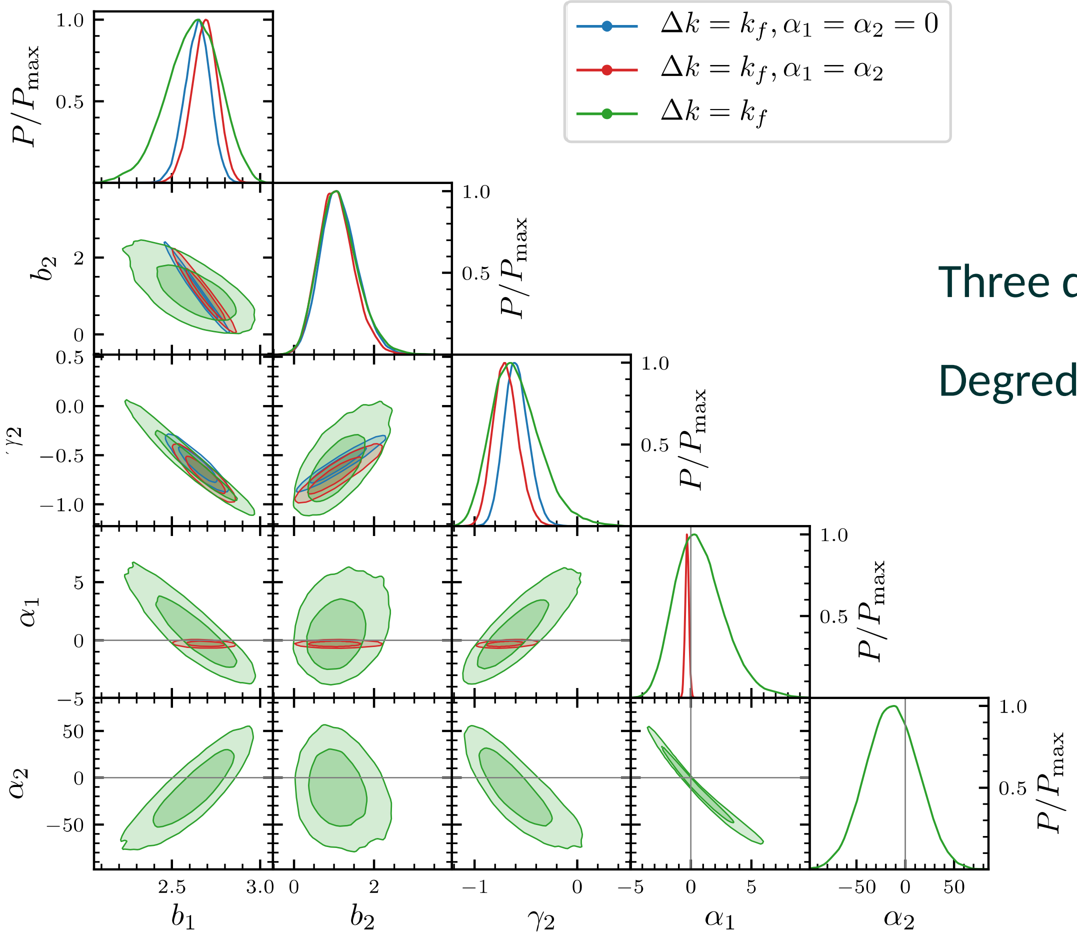
Deviance Information Criterion: $\mathrm{DIC} = \langle D(\boldsymbol{\theta})\rangle_{\mathrm{post}} + \frac{1}{2}\mathrm{Var}(D(\boldsymbol{\theta}))$

The second term (effective number of parameters) penalizes models with many parameters. We are interested in the difference of DIC between two models, $\Delta\mathrm{DIC}$, favoring the model with lower DIC, and with differences of 5 being substantial.

The **Savage-Dickey density ratio** gives the Bayes factor in the case of **properly nested models**:

$$\mathrm{BF}_{jk} = \frac{\int P(\boldsymbol{\theta}_j, \boldsymbol{\psi} = \mathbf{c}|M_k)\,\mathrm{d}\boldsymbol{\theta}_j}{\int \pi(\boldsymbol{\theta}_j, \boldsymbol{\psi} = \mathbf{c}, |M_k)\,\mathrm{d}\boldsymbol{\theta}_j}$$

Model selection: shot-noise parameters



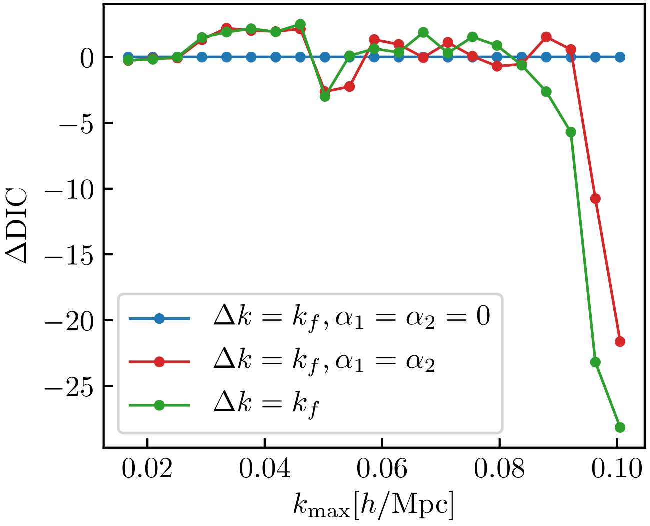
Three different models: $\alpha_1 = \alpha_2 = 0$, $\alpha_1 = \alpha_2$, $\{\alpha_1, \alpha_2\}$

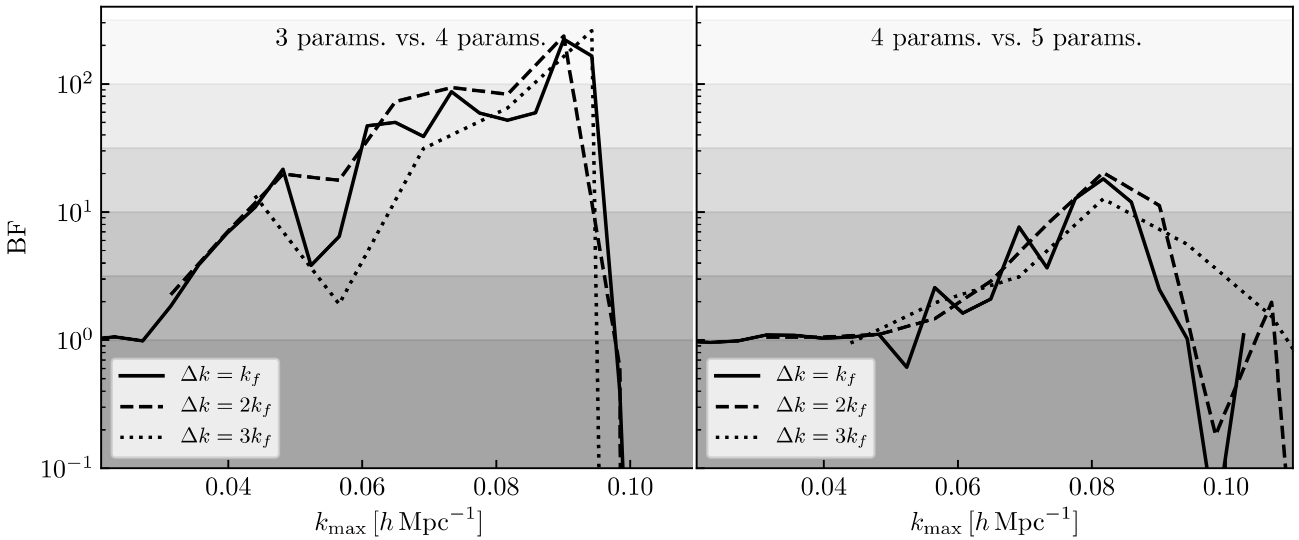
Degredation of posteriors when adding parameters.

[AO+19 in prep.]

Model selection: shot-noise parameters

DIC does not prefer substantially any model only at smaller scales (more important non-linearities), but Bayes Factor favors models with fewer parameters even at larger scales.

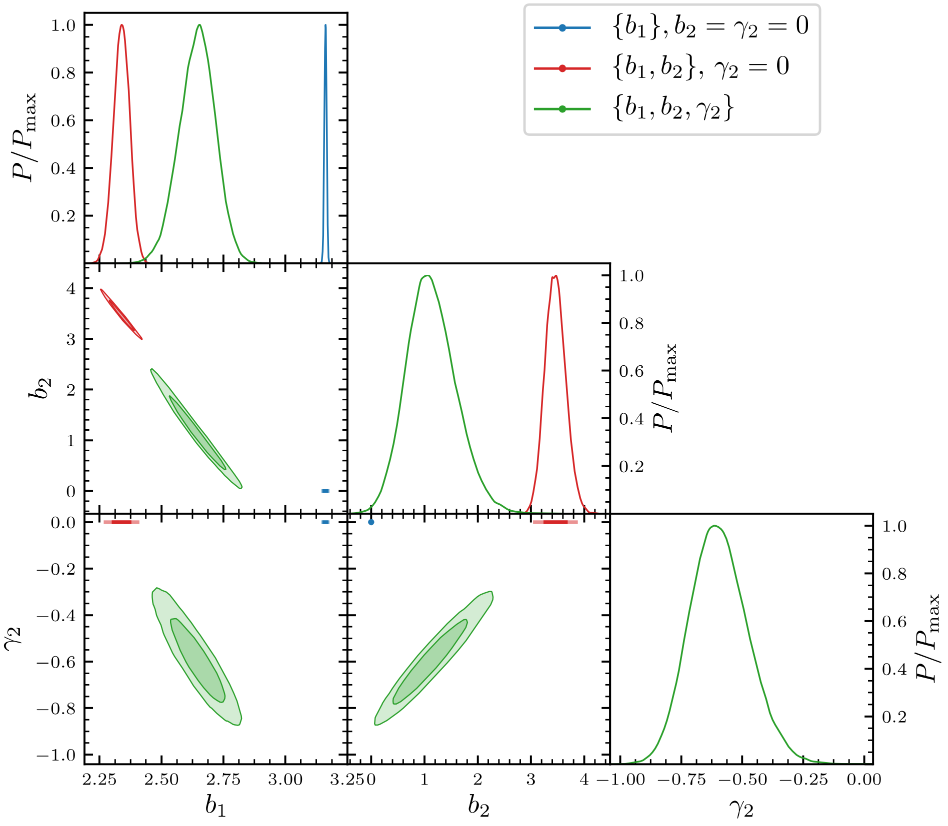


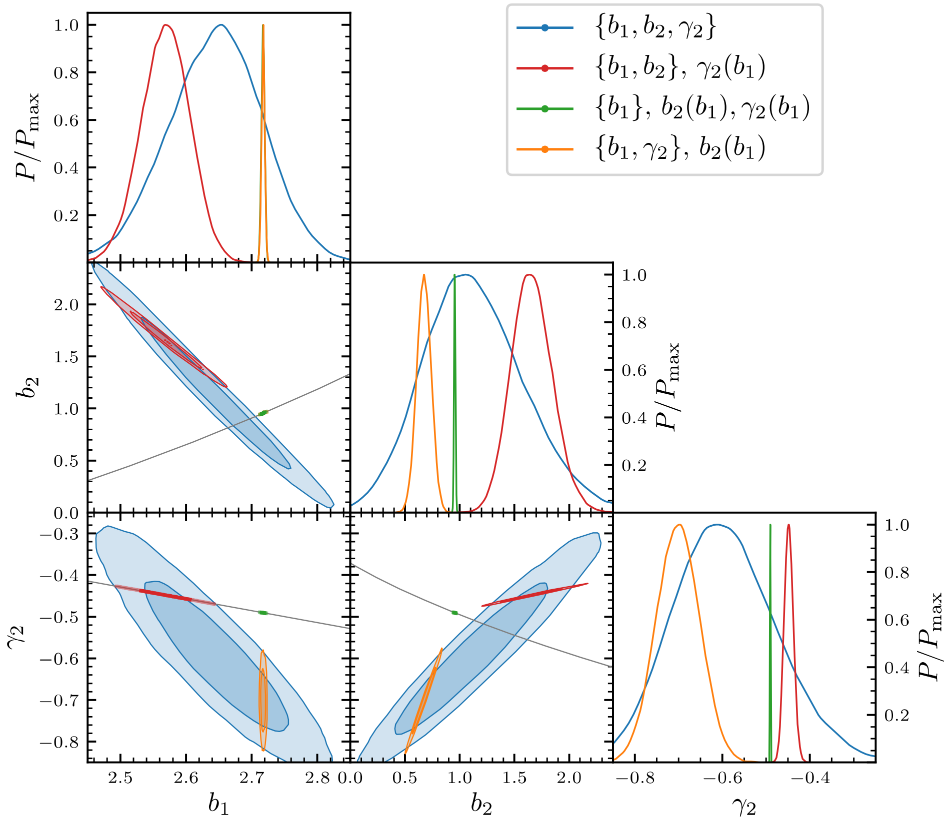


[AO+19 in prep.]

Model selection: bias parameters

We tested some (unphysical) local bias models and a few models relating bias parameters through fitting functions, to lower the dimensionality of parameter space.

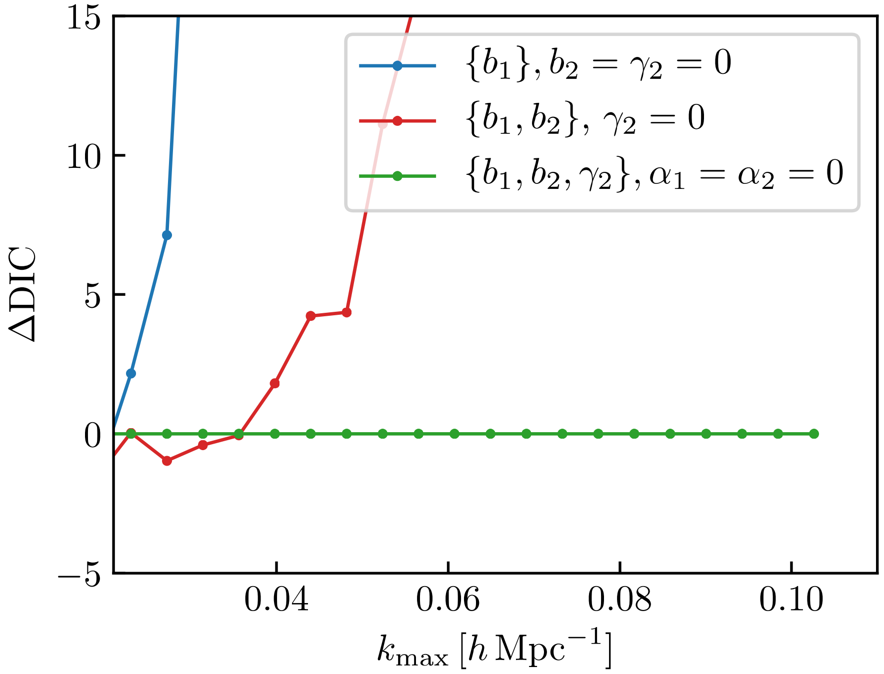


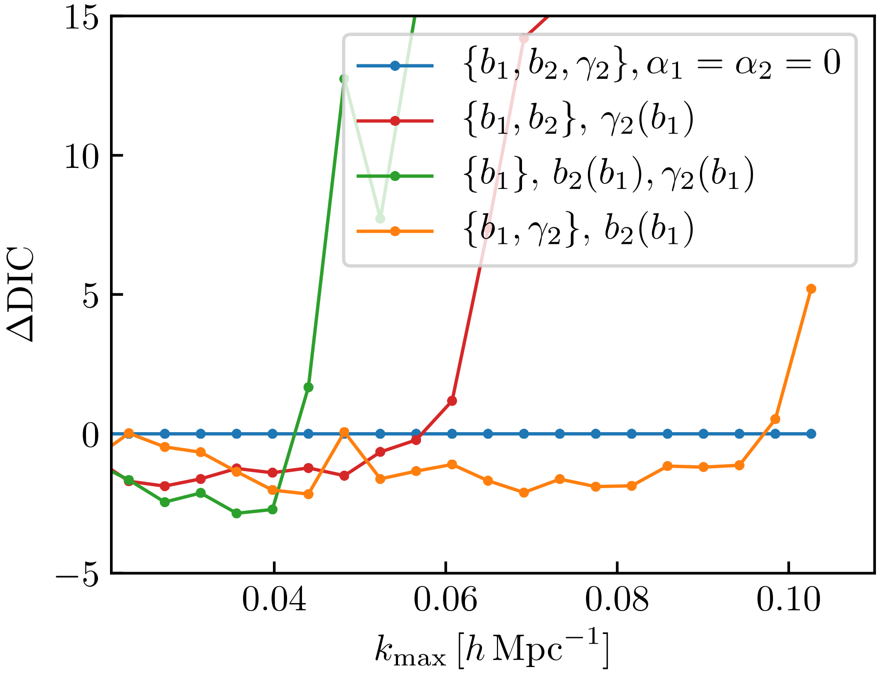


[AO+19 in prep.]

Model selection: bias parameters

The local models are all discarded wrt our “benchmark” model; from the other models, only the one from [Lazeyras *et al.* 2016] works as well as the “benchmark” model, with smaller posteriors.



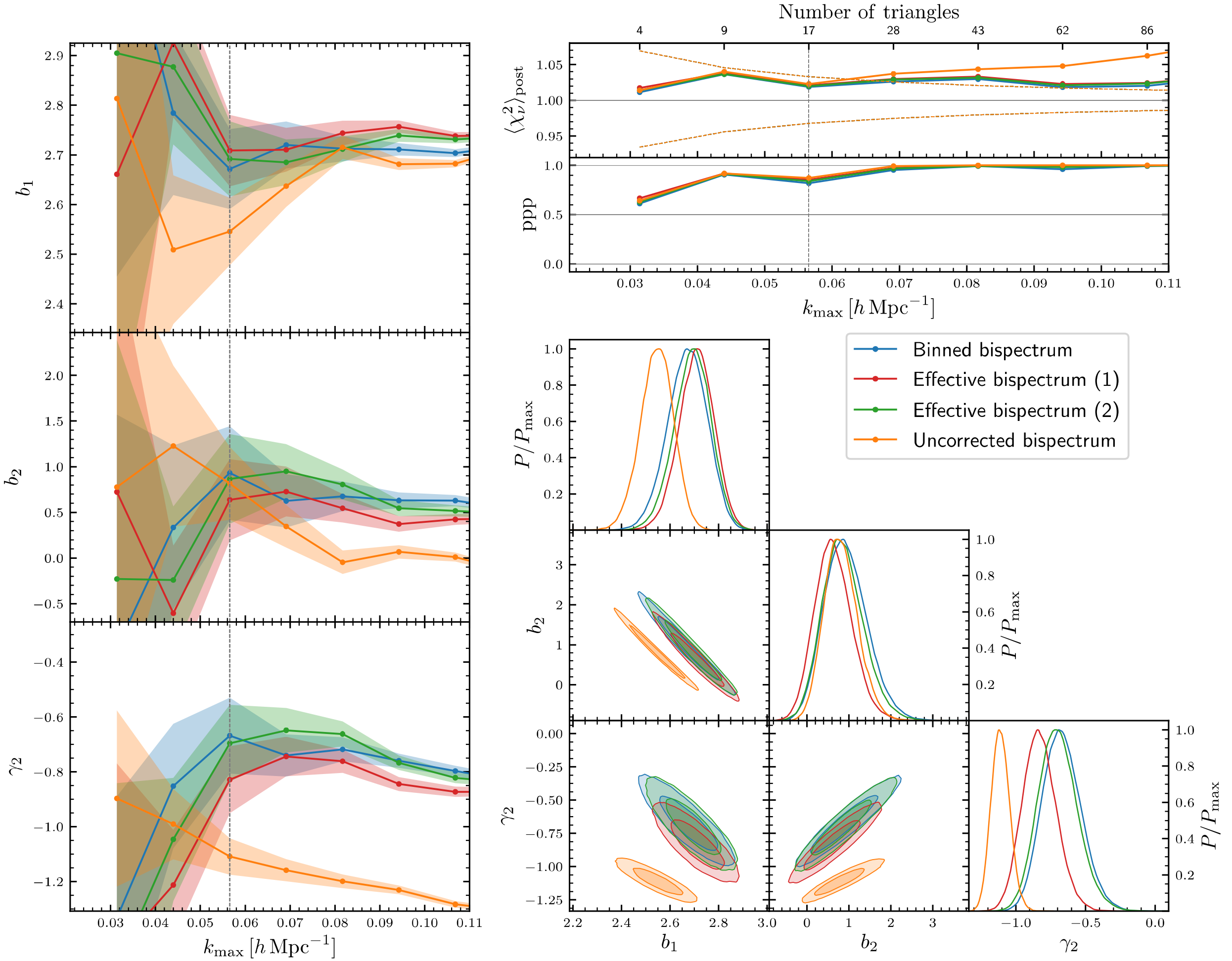


[AO+19 in prep.]

Results: binning effects

Comparison between model properly averaged (binned), evaluated at average modes (effective), or evaluated at central modes (uncorrected).

“Effective” gives similar results wrt to “binned”; “uncorrected” is strongly biased wrt to the binned case, and works the worst.

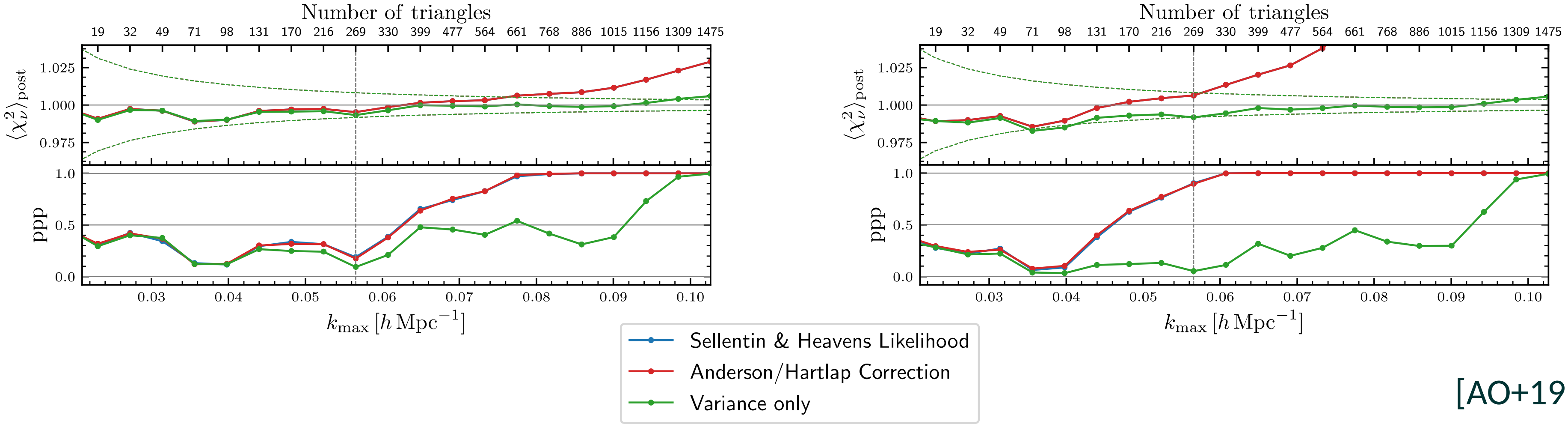


[AO+19 in prep.]

Results: covariance statistical uncertainty

Covariance with 10000 mocks

Covariance with 2000 mocks



[AO+19 in prep.]

- Sellentin & Heavens vs Anderson/Hartlap give the same results.
- Variance-only gives similar parameters posteriors, but less consistent at smaller scales.
- Goodness-of-fit tends to favour the variance-only likelihood, but could be due to the matching of mocks and N-body.
- Worse goodness-of-fit if we use 2000 mocks to evaluate the covariance matrix.