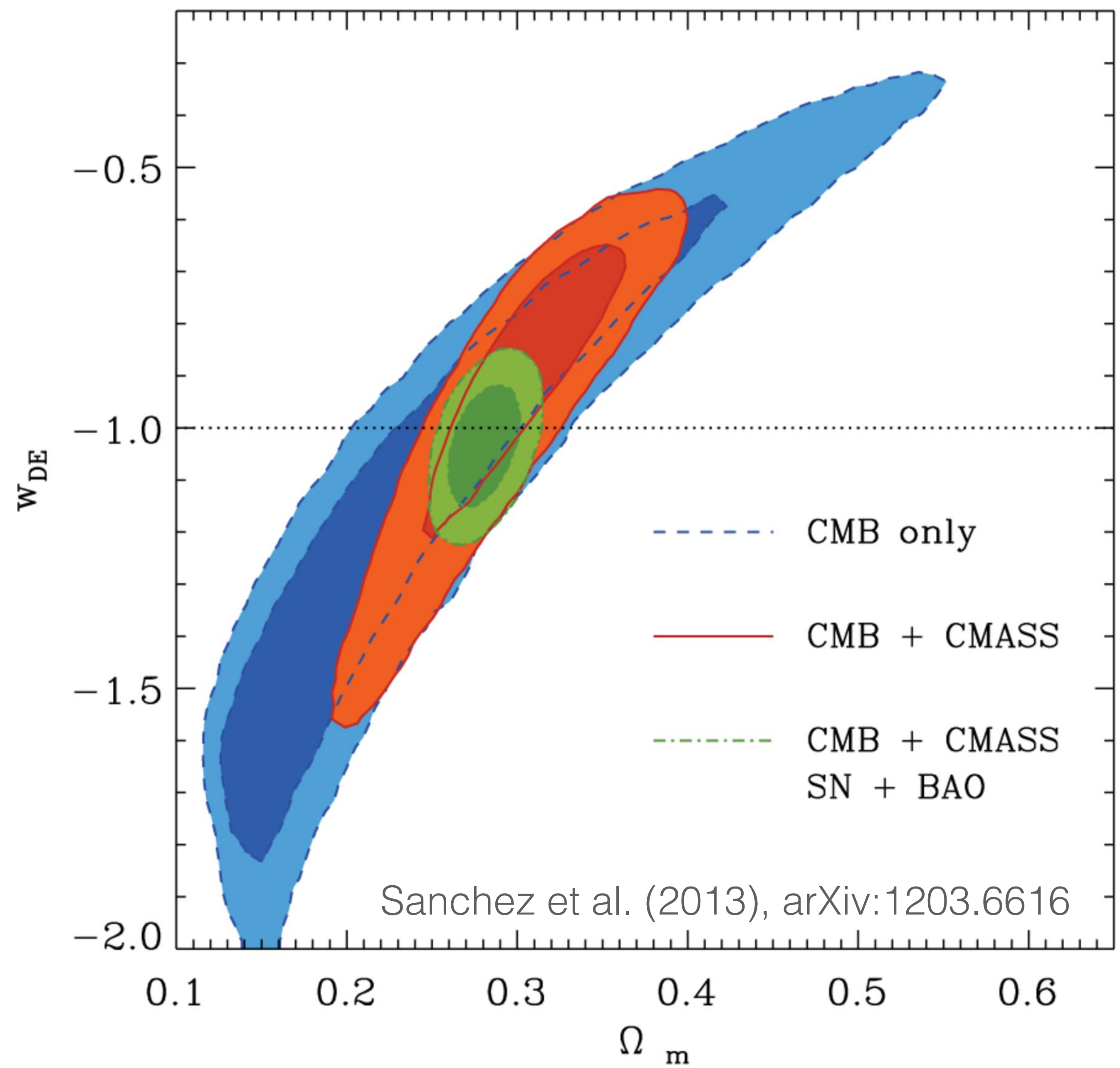
LSST DESC DE School
February 2, 2015

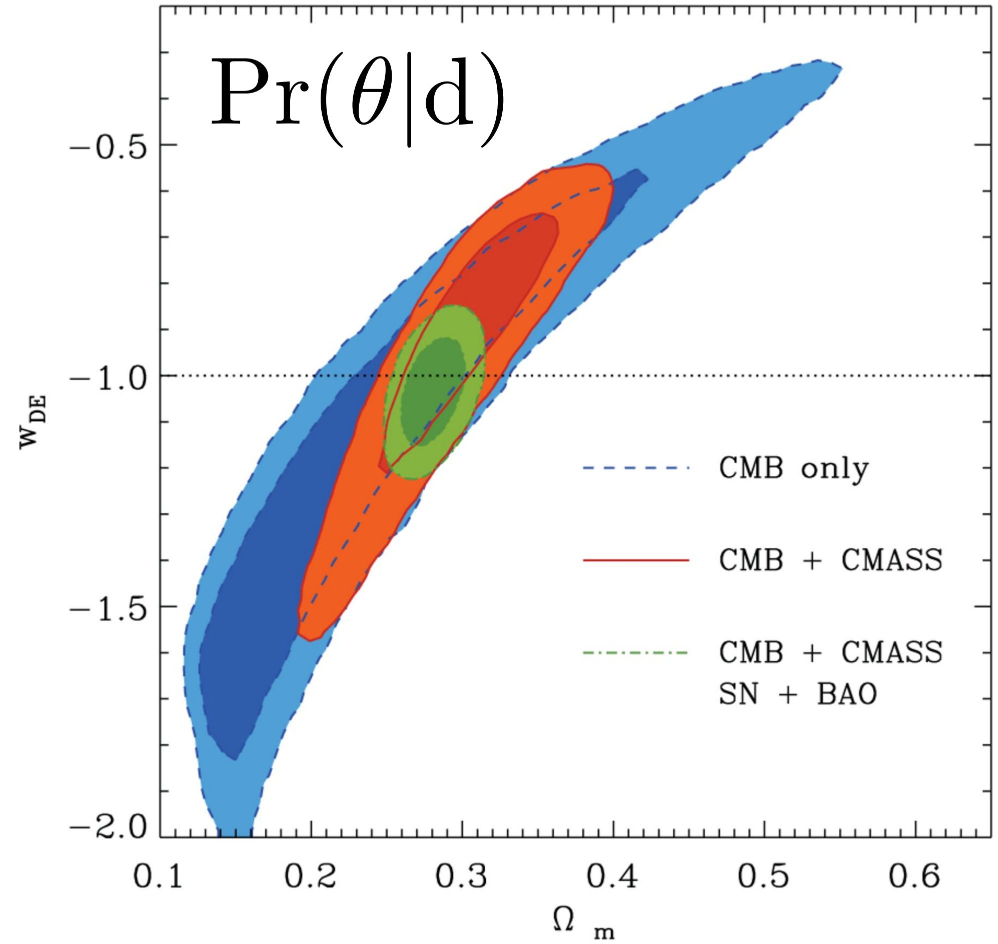
Cosmic (Co)Variance:

How we infer cosmological parameter constraints in the presence of correlated errors

Michael D. Schneider

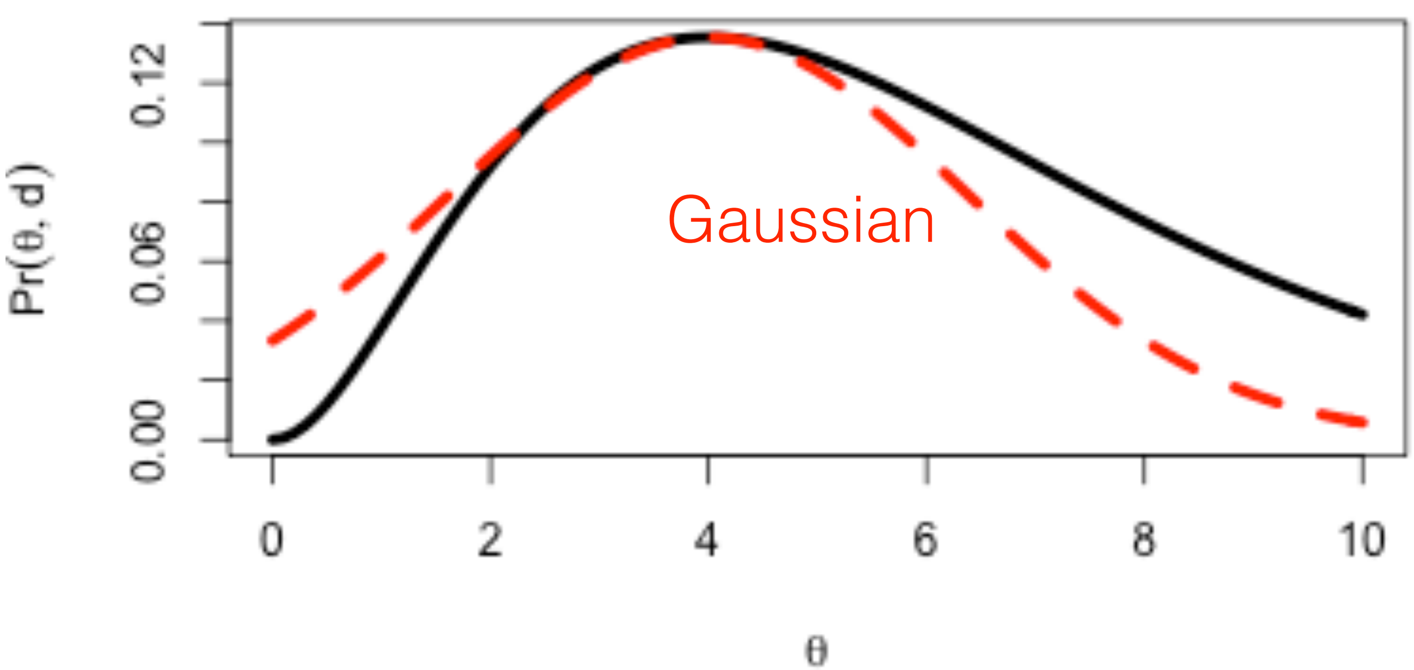


How do we infer cosmological parameters from large-scale structure probes?



These 2D contour plots are showing confidence intervals (e.g., "1-sigma" interval) of the joint posterior probability density of the cosmological parameters given the data.

The posterior is well approximated by a multivariate Gaussian as the parameter constraints get tighter.



So, the mean and covariance of the parameter posterior are the primary quantities to understand expected parameter constraints.

Bayes' theorem terminology

Posterior **Likelihood** **Prior**

$$\mathrm{Pr}(\theta|\mathrm{d}) \propto \mathrm{Pr}(\mathrm{d}|\theta)\,\mathrm{Pr}(\theta)$$

Combining parameter constraints from different surveys

- If 2 surveys (e.g., BOSS and Planck) are statistically independent then their joint likelihood function is the product of the individual likelihoods,

$$\mathrm{Pr}(\mathrm{d}_{\mathrm{BOSS}}, \mathrm{d}_{\mathrm{Planck}}|\theta) = \mathrm{Pr}(\mathrm{d}_{\mathrm{BOSS}}|\theta)\mathrm{Pr}(\mathrm{d}_{\mathrm{Planck}}|\theta)$$

- Assuming Gaussian dependence on the cosmological parameters,

$$\mathrm{Pr}(\mathrm{d}_{\mathrm{BOSS}}, \mathrm{d}_{\mathrm{Planck}}|\theta) \propto G_\theta\left(\Sigma_c\left(C^{-1}_{\mathrm{BOSS}}\hat{\theta}_{\mathrm{BOSS}} + C^{-1}_{\mathrm{Planck}}\hat{\theta}_{\mathrm{Planck}}\right), \Sigma_c\right)$$

$$\Sigma_c \equiv \left(C^{-1}_{\mathrm{BOSS}} + C^{-1}_{\mathrm{Planck}}\right)^{-1}$$

Careful: Flat priors on the cosmological parameters are assumed!

Exercise 1

Given a 2D covariance matrix C for 2 cosmological parameters, what determines the,

1) **area**
2) **orientation**

of the error ellipse?

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Solution 1

Starting with the covariance,

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

our objective is to rotate the coordinate axes so the rotated parameters are uncorrelated,

$$C = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \tilde{\sigma}_1^2 & 0 \\ 0 & \tilde{\sigma}_2^2 \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}$$

Solving the resulting system of equations gives the rotation angle and the semi-major and semi-minor axes of the ellipse, from which we can calculate the area,

$$\mathrm{Area} = \pi\sigma_1\sigma_2\sqrt{1-\rho^2} \qquad \mathrm{Angle} = \frac{1}{2}\tan^{-1}\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}$$

https://magicbeansinc.shinyapps.io/DESC_DESchool2015/error_ellipse.Rmd

See also arXiv:0906.4123

We can estimate the cosmological parameter covariance from the curvature of the likelihood function at its peak,

$$F_{ab} \equiv -\left\langle \frac{\partial^2 \ln L(\mathrm{d}|\theta)}{\partial\theta_a \partial\theta_b} \right\rangle\Bigg|_{\theta=\theta_{\mathrm{ML}}} \approx \mathrm{Cov}^{-1}\mathrm{Pr}(\theta|\mathrm{d})$$

F is called the Fisher matrix and gives a lower bound on the parameter uncertainties for a given:

1. survey
2. data vector
3. likelihood function

This lower bound is called the Cramer-Rao bound.

For cosmological correlation functions ξ, we frequently assume the likelihood is multivariate Gaussian (we'll discuss why later),

$$L(\mathrm{d} \equiv \xi|\theta) \propto |C|^{-1/2} \exp\left(-\frac{1}{2}(\hat{\xi} - \xi(\theta))^T C^{-1} (\hat{\xi} - \xi(\theta))\right)$$

where C is the covariance matrix of the correlation function.

The Fisher matrix can then be reduced to,

$$F_{ab} = \sum_{ij} \frac{\partial \xi_i}{\partial \theta_a} C_{ij}^{-1} \frac{\partial \xi_j}{\partial \theta_b}$$

The data vector covariance C tells us the relative contributions of different angular bins in the correlation function to the cosmological parameter constraints.

For cosmological correlation functions ξ, we frequently assume the likelihood is multivariate Gaussian (we'll discuss why later),

$$L(\mathrm{d} \equiv \xi|\theta) \propto |C|^{-1/2} \exp\left(-\frac{1}{2}(\hat{\xi} - \xi(\theta))^T C^{-1} (\hat{\xi} - \xi(\theta))\right)$$

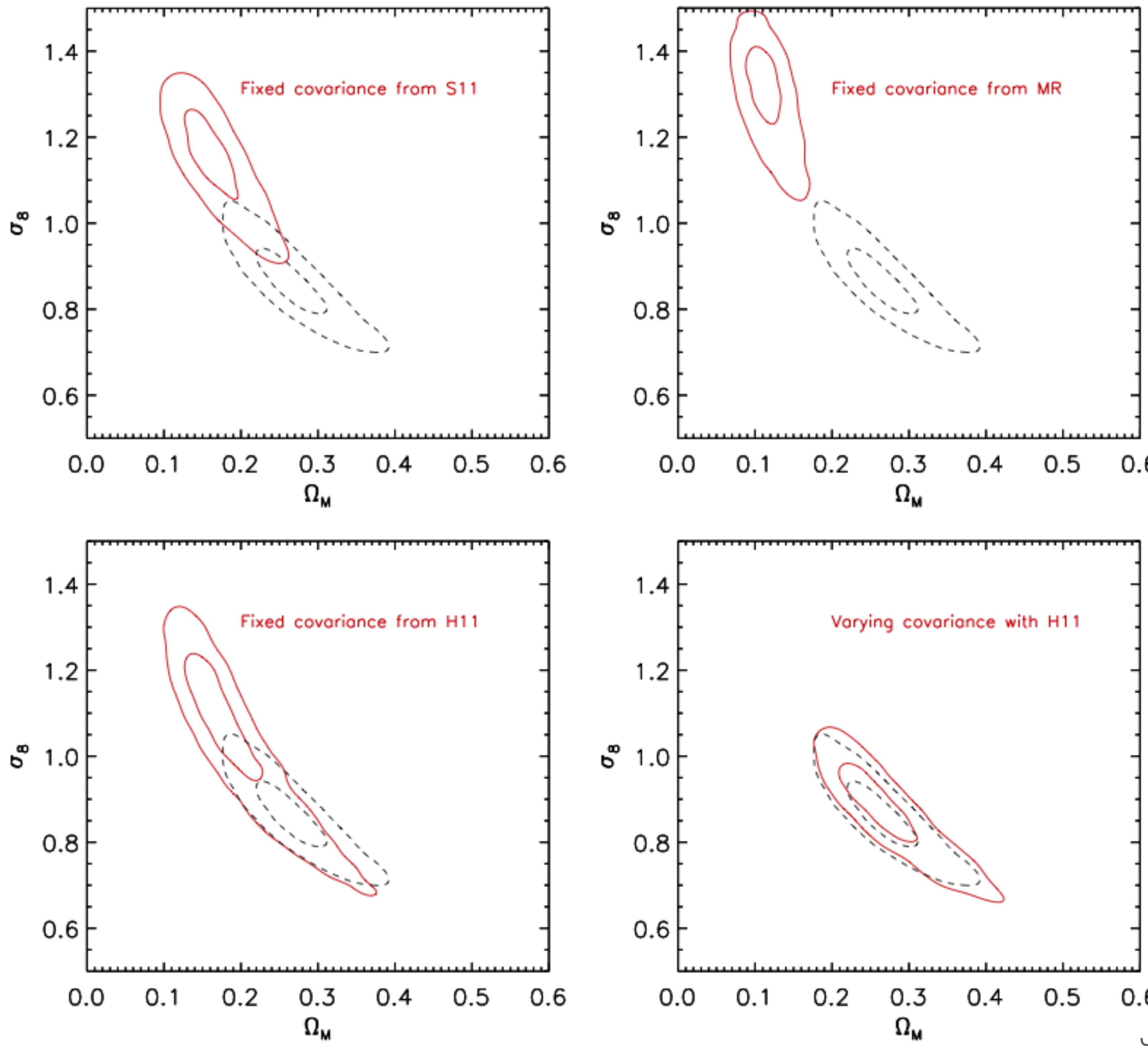
where C is the covariance matrix of the correlation function.

The Fisher matrix can then be reduced to,

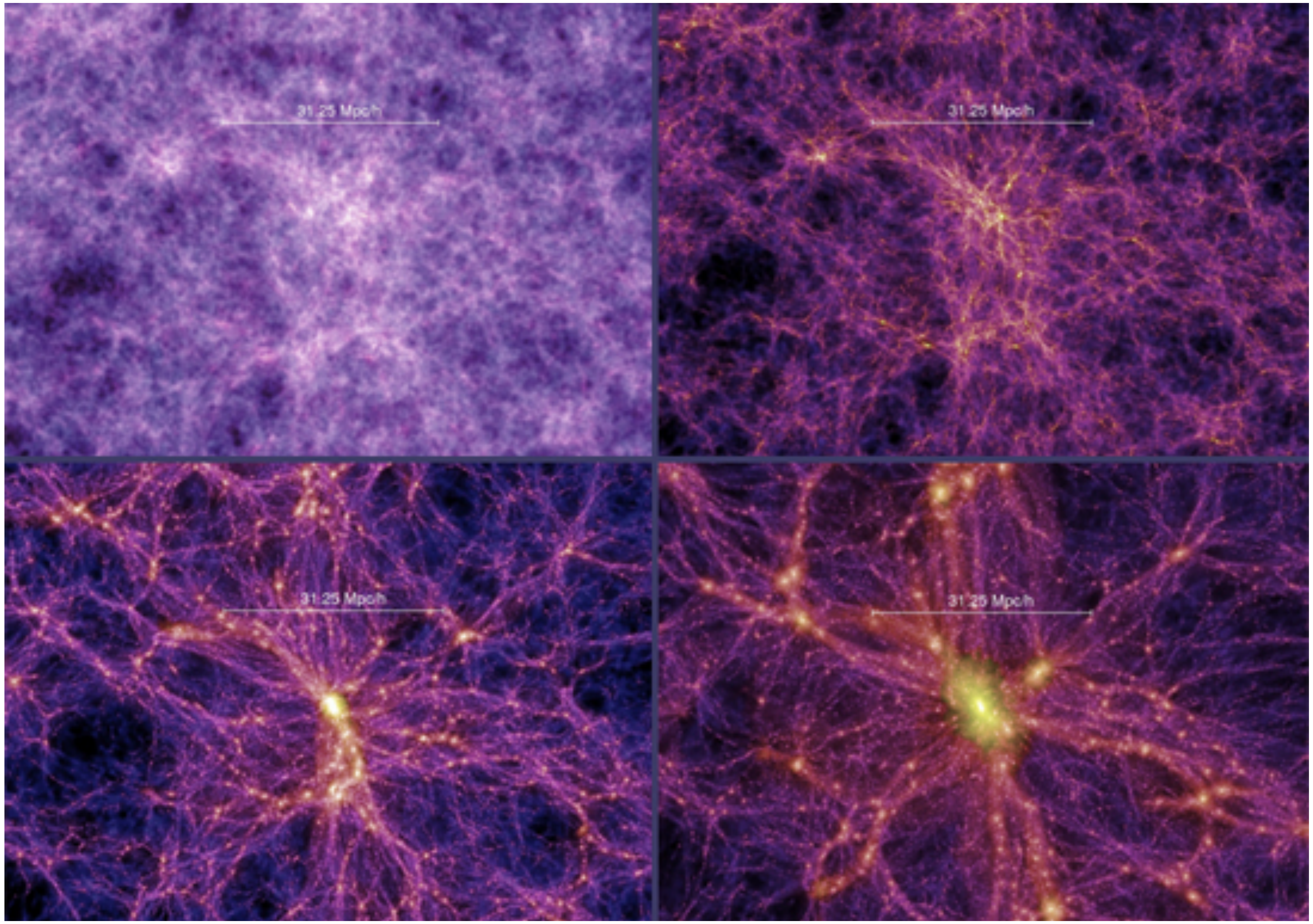
$$F_{ab} = \sum_{ij} \frac{\partial \xi_i}{\partial \theta_a} C^{-1}_{ij} \frac{\partial \xi_j}{\partial \theta_b} + \frac{1}{2} \sum_{ijj'i'} \frac{\partial C_{ij}}{\partial \theta_a} C^{-1}_{jj'} \frac{\partial C_{j'i'}}{\partial \theta_a} C^{-1}_{i'i}$$

The data vector covariance C tells us the relative contributions of different angular bins in the correlation function to the cosmological parameter constraints.

Example: Modeling cosmology dependence of the shear correlation covariance causes a shift along a parameter degeneracy direction in the Deep Lens Survey.



Jee, Tyson, MDS+(2013)



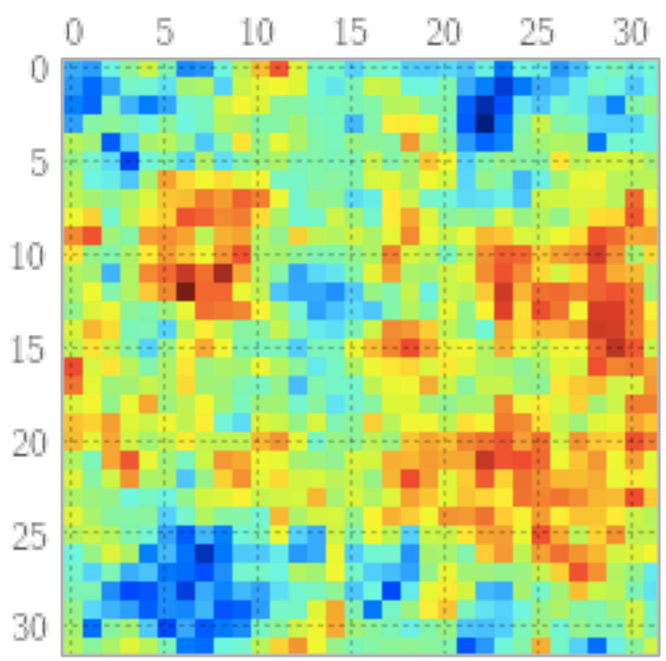
Describing gravitational growth of cosmological matter density perturbations

Predictions of the covariance of summary statistics of the late-time cosmological mass density

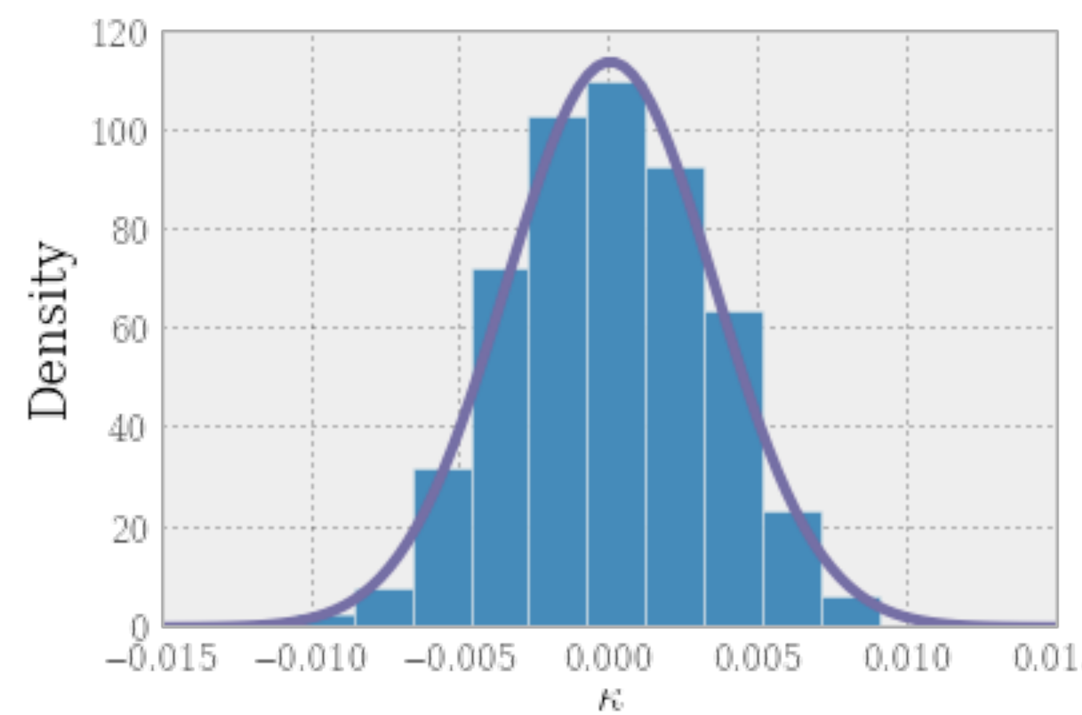
“Vanilla” inflation predicts the initial cosmological mass density perturbations are Gaussian distributed.

What does “Gaussian distributed” mean here?

Bin the mass density into cells



Histogram the cell values



The histogram of cell values is fit well by a Gaussian distribution.

Exercise 2

A) What does the covariance matrix look like for the Fourier transform of a homogeneous Gaussian random field?

B) What does the covariance matrix look like for the power spectrum of a homogeneous and isotropic Gaussian random field?

Solution 2

A)

From Kirkby's lesson:

$$\left\langle \delta(\vec{k})\delta^*(\vec{k}) \right\rangle = (2\pi)^3 \delta_{\rm D}(\vec{k} - \vec{k}')P(\vec{k})$$

B)

(assuming homogeneity)

$$\begin{aligned}
\left\langle \left(\hat{P}(\vec{k}) - P(\vec{k})\right)\left(\hat{P}(\vec{k}') - P(\vec{k}')\right) \right\rangle &= \left\langle \hat{P}(\vec{k})\hat{P}(\vec{k}') \right\rangle - P(k)P(k') \\
&= \left\langle \delta(\vec{k})\delta^*(\vec{k})\delta(\vec{k}')\delta^*(\vec{k}') \right\rangle - P(k)P(k') \\
&= \left\langle \delta(\vec{k})\delta^*(\vec{k}) \right\rangle \left\langle \delta(\vec{k}')\delta^*(\vec{k}') \right\rangle + \left\langle \delta(\vec{k})\delta(\vec{k}) \right\rangle \left\langle \delta^*(\vec{k})\delta^*(\vec{k}') \right\rangle + \\
&+ \left\langle \delta(\vec{k})\delta^*(\vec{k}') \right\rangle \left\langle \delta^*(\vec{k})\delta(\vec{k}') \right\rangle - P(k)P(k') \\
&= P(k)^2(2\pi)^3 \left(\delta_{\rm D}(\vec{k} - \vec{k}') + \delta_{\rm D}(\vec{k} + \vec{k}')\right)
\end{aligned}$$

So the covariance of the power spectrum is the square of the power spectrum!

Assuming $\left\langle \delta(\vec{k})\delta^*(\vec{k})\delta(\vec{k}')\delta^*(\vec{k}') \right\rangle_c = 0$

The matter power spectrum covariance after inflation…

- Is diagonal
  - Every Fourier mode of the mass density and the mass density power spectrum is statistically independent.
  - Consequence of homogeneity
- Depends only on the power spectrum
  - General result for a Gaussian random field.

The power spectrum is the theorist's favorite summary statistic in part because of these properties.

The shell-averaged power spectrum estimator

- Assuming isotropy, power spectrum estimates with the same wave vector modulus but different phase are independent estimators of the same band power.

- Define the k "shell-averaged" estimator,

$$\hat{P}(k) \equiv \frac{1}{N_k} \sum_{i=1}^{N_k} \hat{P}(\vec{k}) \qquad\qquad N_k \equiv 4\pi k^2 \delta k V$$

- The covariance becomes,

$$\mathrm{Cov}(\hat{P}(k)) = \frac{2}{N_k} P^2(k)$$

Growth of mass density mode correlations…

…is caused by gravitational collapse, which breaks homogeneity

The covariance of the power spectrum depends on the 4-point expectation, $\left\langle \delta(\vec{k})\delta^*(\vec{k})\delta(\vec{k}')\delta^*(\vec{k}') \right\rangle$

If the joint probability distribution of the 4 δ terms does not factor, then there is a "connected" 4-point function that also contributes to the covariance,

$$\left\langle \delta(\vec{k})\delta^*(\vec{k})\delta(\vec{k}')\delta^*(\vec{k}') \right\rangle_c \equiv T(\vec{k}, -\vec{k}, \vec{k}', -\vec{k}')$$

The trispectrum T is nonzero only when the density field becomes non-Gaussian through gravitational evolution.

The nonlinear power spectrum covariance

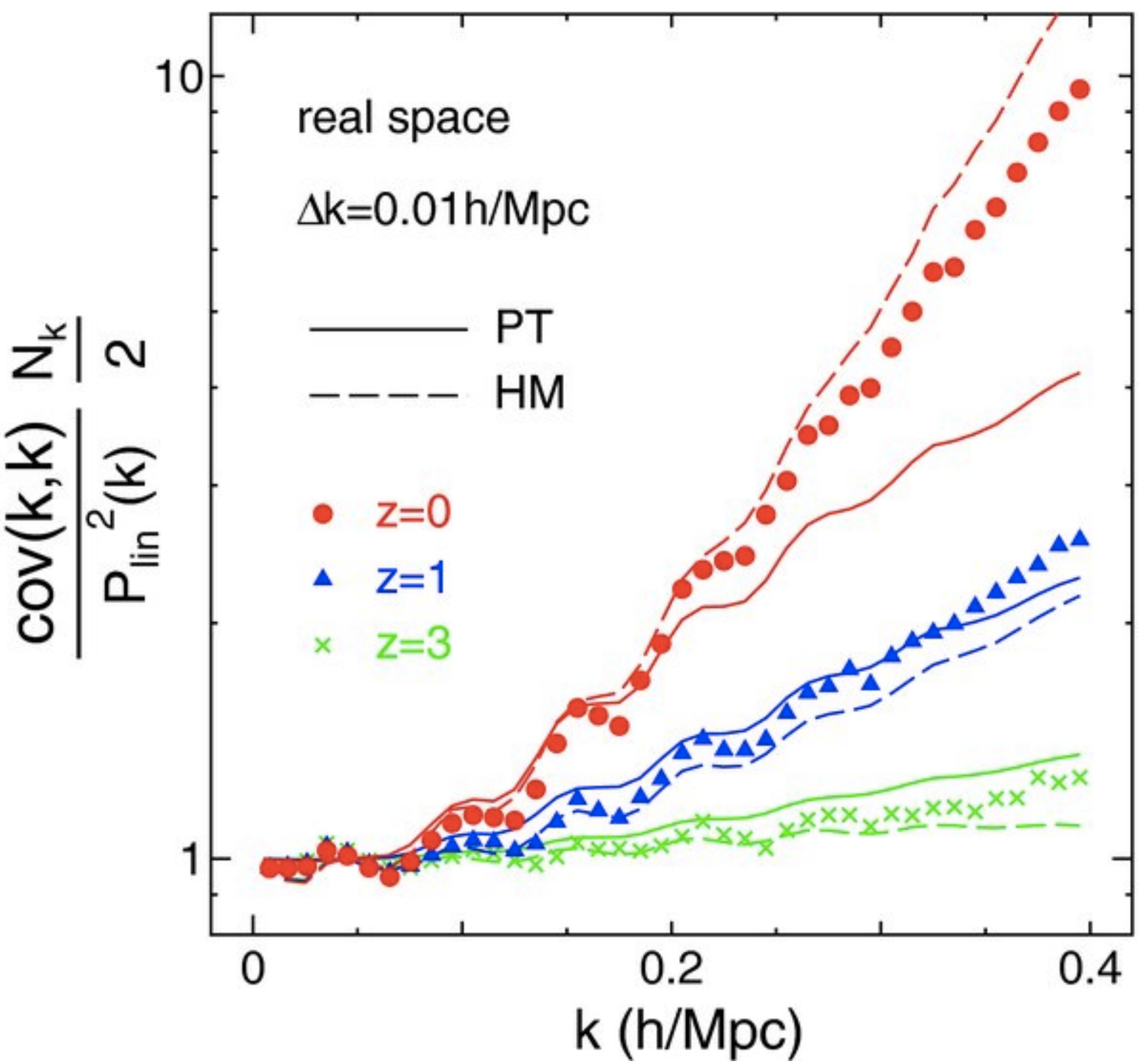
- The non-Gaussianity of the mass density field admits a non-zero trispectrum, which leads to off-diagonal terms in the power spectrum covariance.

$$C_{ij} = \frac{2}{N_k} P_i^2 \delta_{ij} + \frac{1}{V} \bar{T}(k_i, k_j) \qquad N_k \equiv 4\pi k^2 \delta k V$$

- Increasing survey volume reduces the power spectrum covariance. But, at fixed V, the trispectrum dominates for large k where there are many modes.

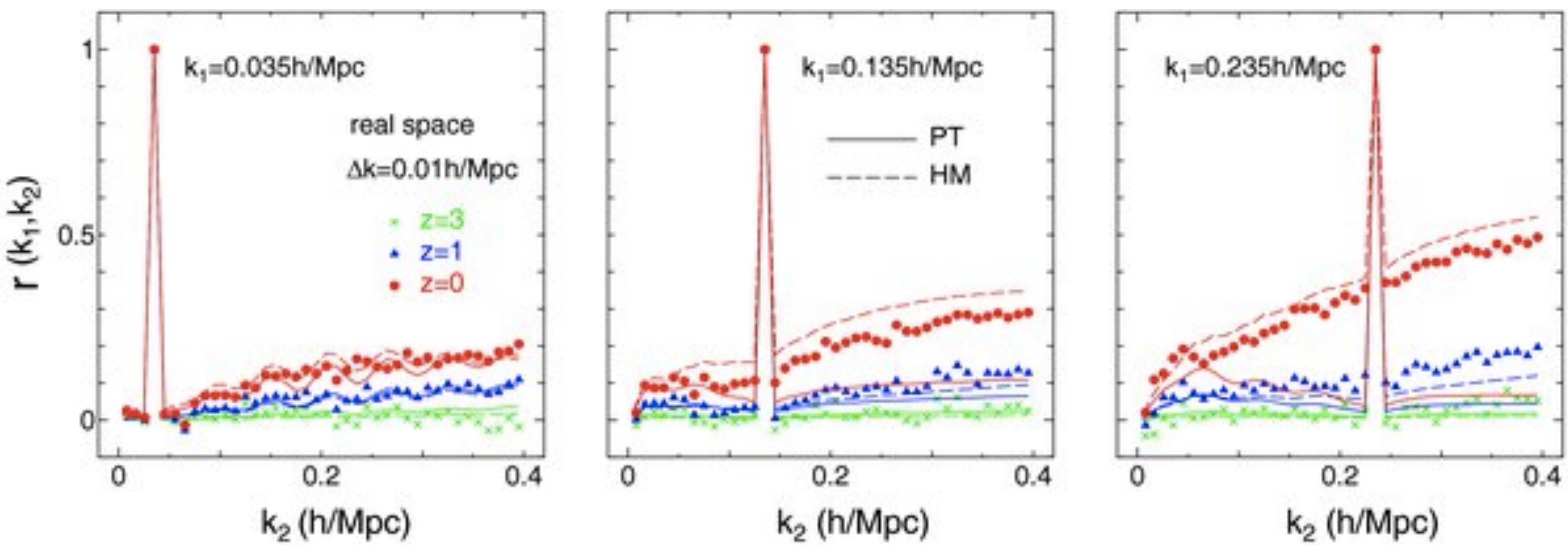
Warning: the shell-averaged trispectrum above neglects an important term - ‘super-sample covariance’.

Growth of P(k) correlations



Takahashi et al. (2009)

Growth of P(k) correlations



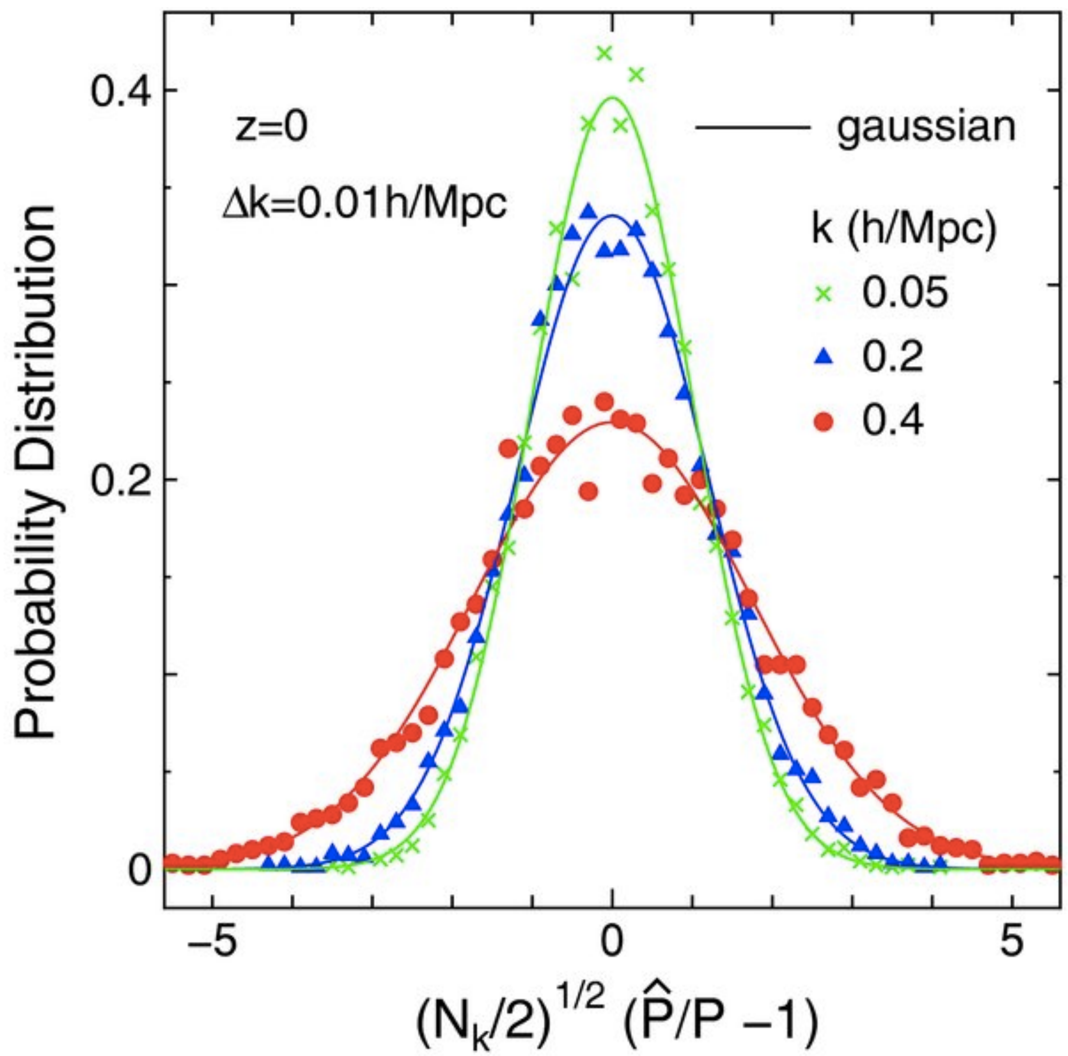
Takahashi et al. (2009)

Gaussian likelihood approximation for P(k)

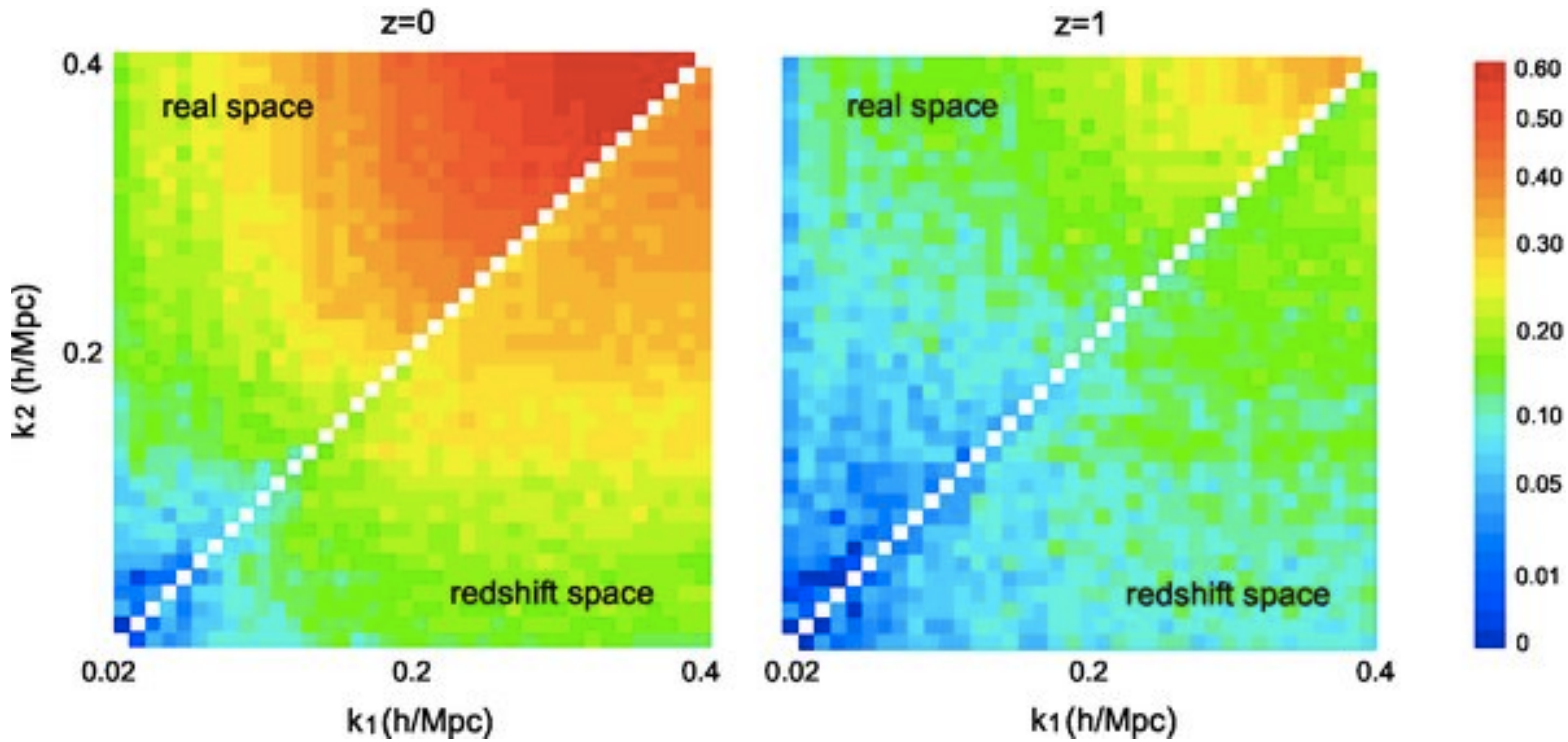
The non-Gaussian mass density field leads to correlations in the band powers P(k).

But, the shell-averaged *estimator*, of P(k) is a sum of many different modes.

It may be reasonably modeled as Gaussian distributed because of the Central Limit theorem.



Takahashi et al. (2009)



Covariances for cosmological data analysis

Exercise 3

List the main steps in a computational algorithm to calculate the galaxy power spectrum covariance matrix from N-body simulations.

Solution 3

Calculating the galaxy power spectrum covariance matrix

1. Generate a realization of the initial mass density on a cubic grid given the linear power spectrum, the seed for the pseudo-random number generator, and a chosen periodic volume and grid spacing.

2. Perturb a set of mass tracer particles using displacements and velocities calculated from the initial mass density grid.

3. Evolve the *N*-body simulation to the present day, saving 'snapshots' of the particle distribution at regular intervals.

4. Run a halo finding algorithm on the save snapshots to identify gravitationally bound halos that will host model galaxies. Build merger trees of the halo histories.

5. Construct the mass and halo distribution along an observable light cone with a specified opening angle by combining and interpolating snapshots (and possibly distinct simulation volumes).

6. Add galaxies to the light cone (via 'halo occupation distribution' or 'semi-analytic' models).

7. Apply the survey window function to the light cone.

8. **Repeat steps 1-7 to obtain N_r samples of the simulated galaxy distribution.**

9. Calculate power spectra and construct the sample covariance estimator from the N_r samples.

Aside: estimating the covariance by sub-sampling the observed area is bad because,

1) the sub-volumes are not independent,
2) the window function is wrong.

Here 'bad' means that the covariance estimator is necessarily biased when derived from the survey (and usually under-predicts the variance).

Using simulations to estimate the covariance can be unbiased in principle.

The sample covariance estimator

Given N_r independent samples of a survey simulation,

$$L\left(\left\{\hat{P}_i\right\}_{i=1}^{N_r}|\theta\right) = \prod_{i=1}^{N_r} L(\hat{P}_i|\theta)$$

$$\propto \prod_{i=1}^{N_r} |C|^{-1/2} \exp\left(-\frac{1}{2}\left(\hat{P}_i - \bar{P}\right)^T C^{-1}\left(\hat{P}_i - \bar{P}\right)\right)$$

$$= |C|^{-N_r/2} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(C^{-1}\sum_{i=1}^{N_r}\left(\hat{P}_i - \bar{P}\right)\left(\hat{P}_i - \bar{P}\right)^T\right)\right)$$

This suggests defining,

$$\hat{C}_{ab} \equiv \frac{1}{N_r}\sum_{i=1}^{N_r}\left(\hat{P}_i(k_a) - \bar{P}(k_a)\right)\left(\hat{P}_i(k_b) - \bar{P}(k_b)\right)$$

In fact, this is the maximum-likelihood estimator for the covariance C.

The sample covariance estimator

What are the properties of the sample covariance estimator? Let's derive the sampling distribution from the likelihood we already have.

$$L\left(\hat{P}|\theta, C\right) d\hat{P} \propto |C|^{-N_r/2} \exp\left(-\frac{N_r}{2}\mathrm{Tr}\left(C^{-1}\hat{C}\right)\right)\left|\frac{d\hat{P}}{d\hat{C}}\right| d\hat{C}$$

To further transform this expression to the sampling distribution for the covariance estimator, we need to

1) replace the mean P in the definition of the C estimator with the sample mean, and
2) calculate the Jacobian.

These steps involve a long bit of math (see Wishart 1928 for the original derivation) to get a standard result:

$$\Pr\left(\hat{C}|\hat{P}, C\right) \propto \left|\hat{C}\right|^{(N_r - N_b - 1)/2} |C|^{-N_r/2} \exp\left(-\frac{N_r}{2}\mathrm{Tr}\hat{C}C^{-1}\right)$$

This is the Wishart distribution.

The sample covariance estimator

The covariance of the sample covariance estimator is given by the second moment of the Wishart distribution,

$$\left\langle \Delta\hat{C}_{ij}\Delta\hat{C}_{mn}\right\rangle = \frac{1}{N_r-1}\left(C_{im}C_{jn}+C_{in}C_{jm}\right) \rightarrow \mathrm{Var}\left(\hat{C}_{ii}\right) = \frac{2}{N_r-1}C_{ii}^2$$

So the variance of the sample covariance is (again) the square of the sample covariance.

This is a nice result, but 2 questions remain:

1) What is the error on the inverse covariance matrix estimator?

2) What level of covariance estimator error can we tolerate in a given analysis?

Taylor, Joachimi, Kitching (2013)

The sample inverse covariance estimator

We can define an estimator for the inverse covariance by inverting our covariance estimator,

$$\hat{\Psi} \equiv \frac{N_r - 1}{N_r} \hat{C}^{-1}$$

This estimator is inverse-Wishart distributed (by definition) and has a known bias,

$$\left\langle \hat{\Psi} \right\rangle = \frac{N_r - 1}{N_r - N_b - 2} \Psi$$

which we can correct for by inverting the known multiplicative factor.

The covariance of the inverse covariance estimator is given by the 2nd moment of the inverse-Wishart distribution,

$$\left\langle \Delta\hat{\Psi}_{ij} \Delta\hat{\Psi}_{mn} \right\rangle = A \left[2\Psi_{ij}\Psi_{mn} + (N_r - N_b - 2)\left(\Psi_{im}\Psi_{jn} + \Psi_{in}\Psi_{jm}\right)\right]$$

$$A \equiv \frac{(N_r - 1)^2}{(N_r - N_b - 1)(N_r - N_b - 2)^2 (N_r - N_b - 4)}$$

Taylor, Joachimi, Kitching (2013)

Propagating covariance estimator errors

Recall the Fisher matrix relating the data covariance to the cosmological parameter covariance is,

$$F_{\alpha\beta} \equiv \frac{\partial \mathrm{d}}{\partial \theta_a} C^{-1} \frac{\partial \mathrm{d}}{\partial \theta_b}$$

If we replace the data covariance with the *estimator* for the covariance, we get a perturbation to the Fisher matrix depending on the variance of the estimator. In general, we may then forecast the cosmological parameter covariance as,

$$\hat{C}^{\theta}_{\alpha\beta} = (F + \Delta F)^{-1}_{\alpha\beta}$$

Plugging in the formula for the error in the inverse covariance estimator from the previous slide,

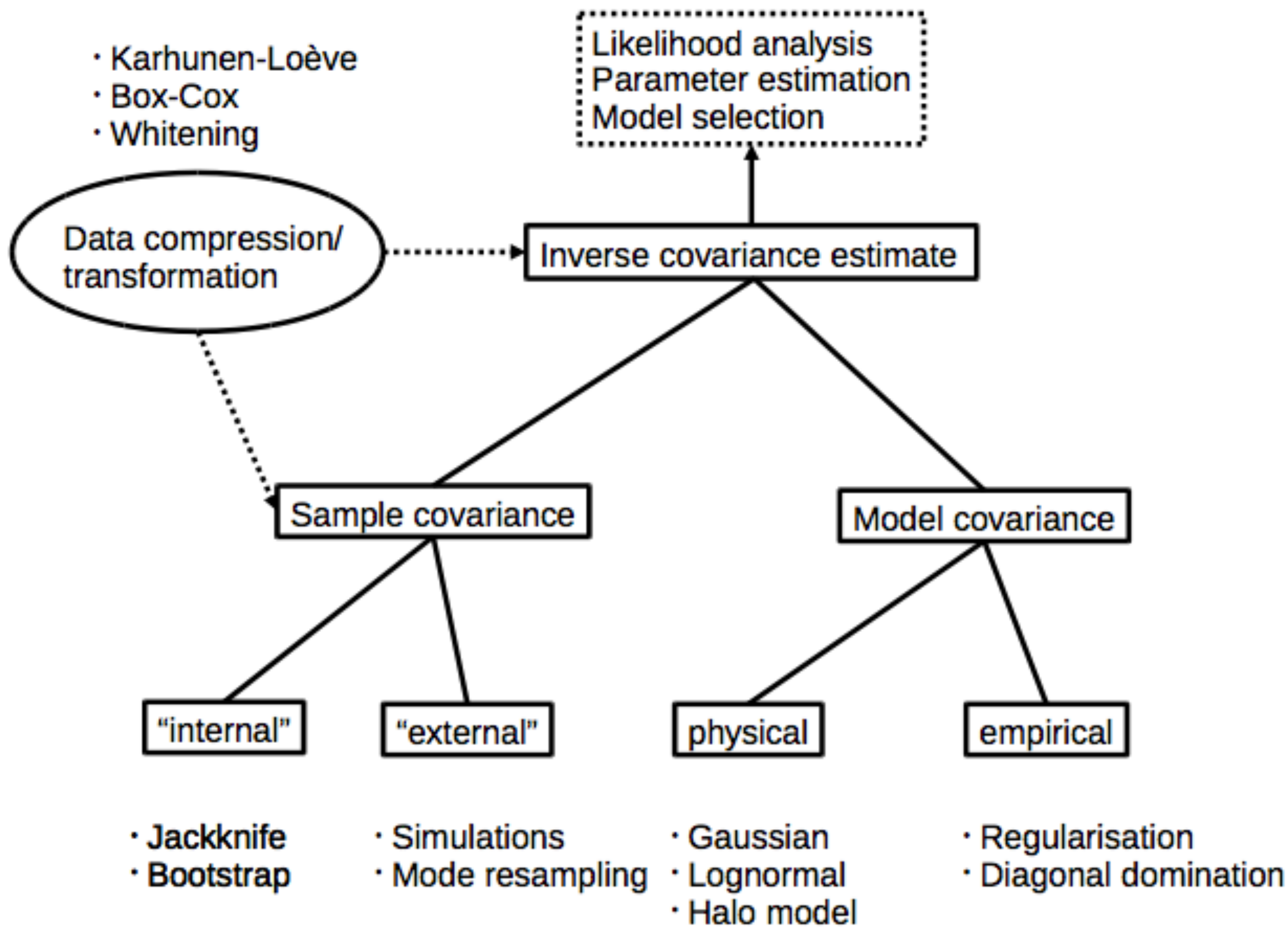
$$\left\langle \Delta \hat{C}^{\theta}_{\alpha\beta} \right\rangle \approx \frac{N_b - N_p}{N_r - N_b} C^{\theta}_{\alpha\beta} \approx \left(1 + \frac{N_b}{N_r}\right) C^{\theta}_{\alpha\beta}$$

So, an error in the data covariance leads to an increase the cosmological parameter errors.

We can use this formula to set bounds on the tolerable error in the data covariance estimators.

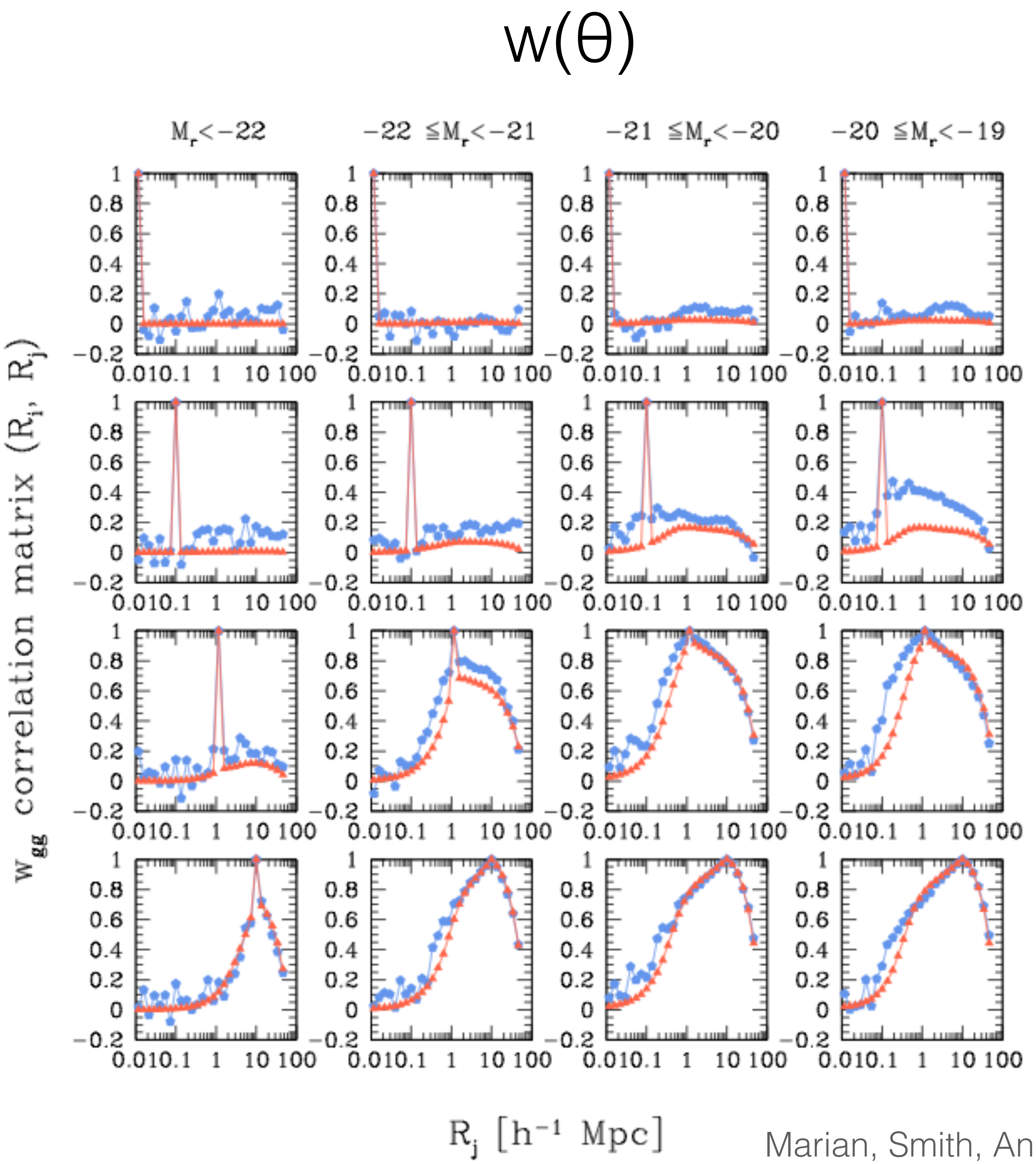
Taylor, Joachimi, Kitching (2013)
Dodelson & MDS (2013)

Various approaches to covariance estimation



Taylor, Joachimi, Kitching (2013)

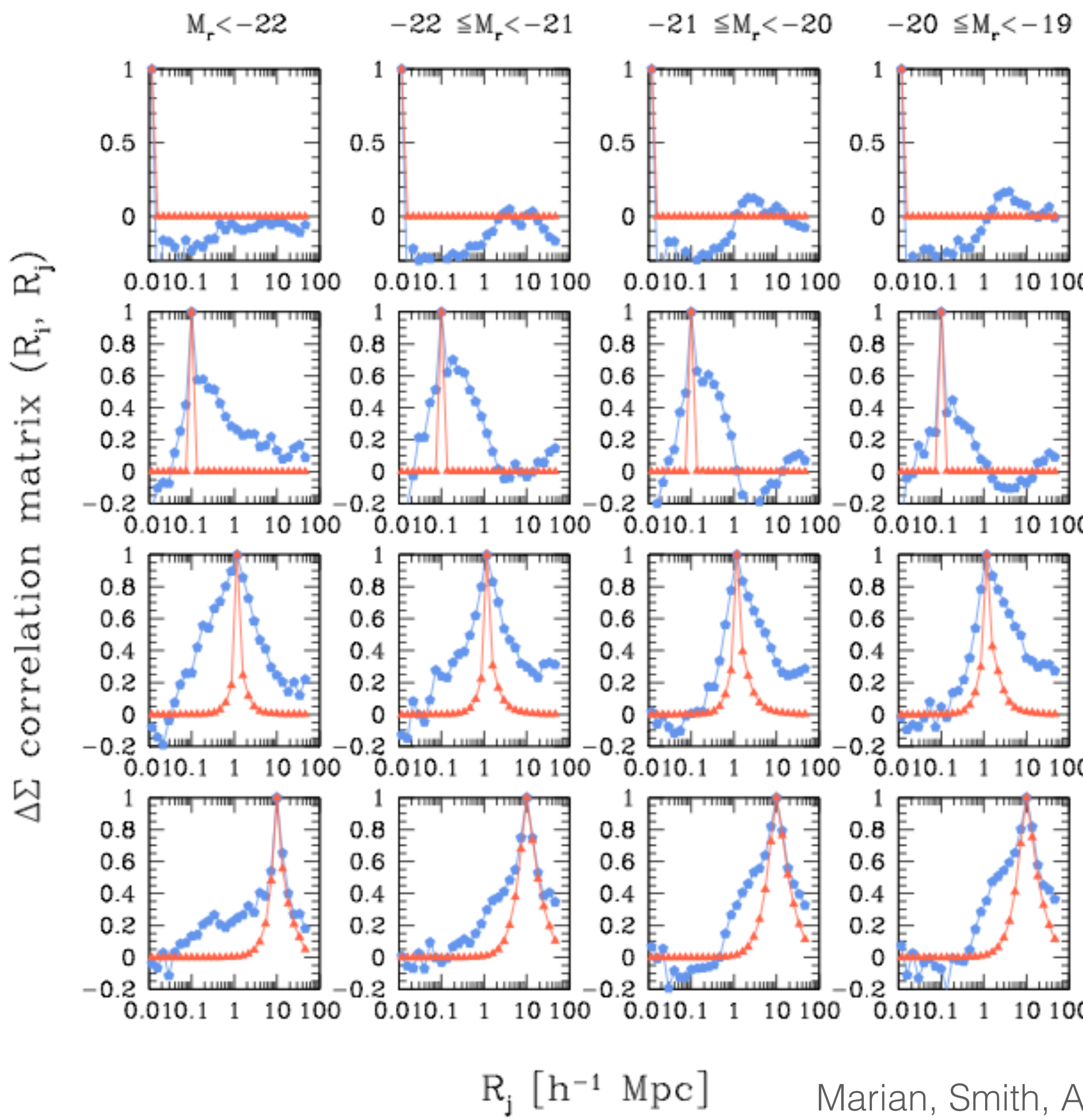
Example: GGL + w(θ) covariance



Marian, Smith, Angulo (2014), arXiv:1410.3468

Example: GGL + w(θ) covariance

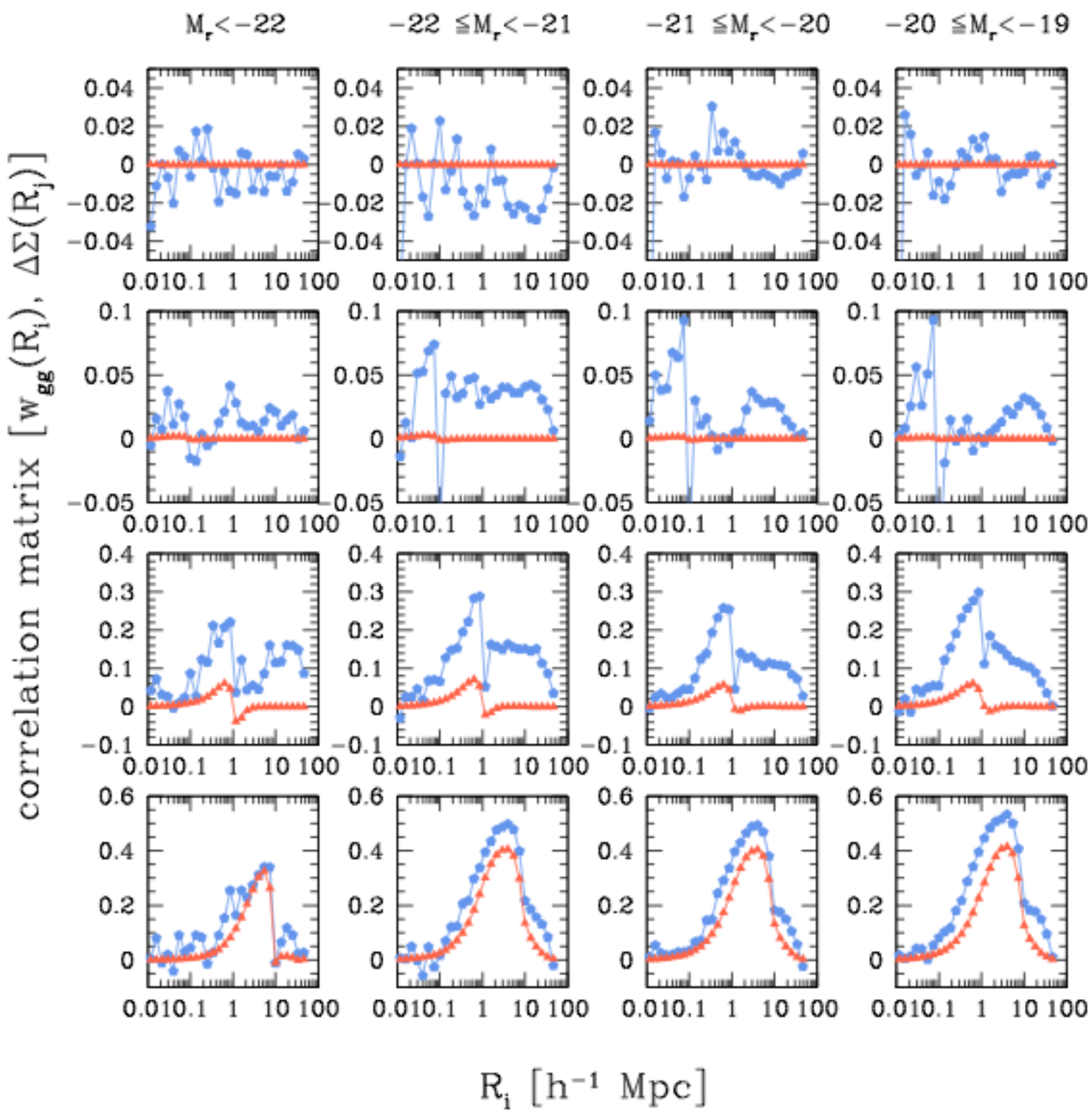
tangential shear



Marian, Smith, Angulo (2014), arXiv:1410.3468

Example: GGL + w(θ) covariance

tangential shear / w(θ) cross-correlation



Marian, Smith, Angulo (2014), arXiv:1410.3468

Homework question

- What is an "optimal estimator" for the clustering of the cosmological mass density field?
  - How does this depend on the covariance of the mass density summary statistic(s)?
  - How does this depend on the functional form of the likelihood?