A + B = C: Estimation of the inverse covariance matrix made easy with *Precision Matrix Expansion*

KIPAC Tea, 05/30/17
Oliver Friedrich
(LMU gravitational lensing; Stella Seitz, Tamas Varga, Matthias Kluge++)

Precision matrix expansion – efficient use of numerical simulations in estimating errors on cosmological parameters

Oliver Friedrich[1,2] ⋆, Tim Eifler[3,4] †

[1] *Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany*
[2] *Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*
[3] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[4] *Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*

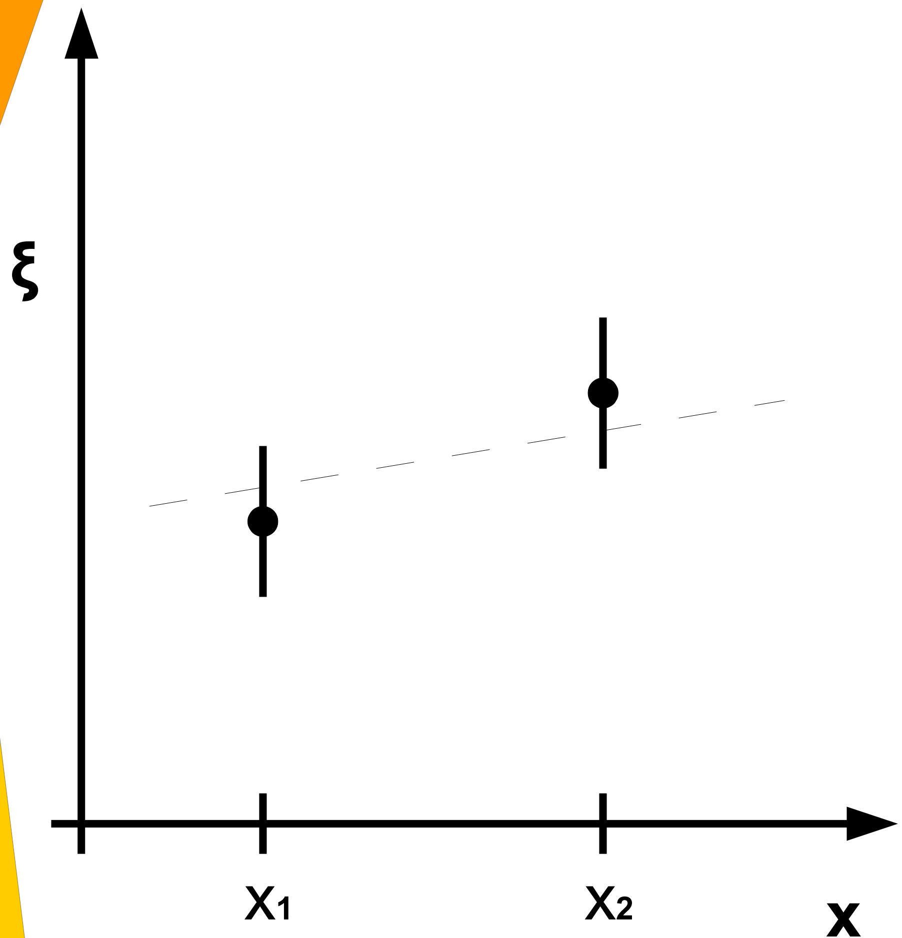
arXiv:1703.07786

A + B = C: Estimation of the inverse covariance matrix made easy with *Precision Matrix Expansion*

- Data vectors & covariance matrices
- Multi-probe analyses
  (new orders of magnitude for covariance estimation)
- Precision matrix expansion

A + B = C: Estimation of the inverse covariance matrix made easy with *Precision Matrix Expansion*

- Data vectors & covariance matrices
- Multi-probe analyses
  (new orders of magnitude for covariance estimation)
- Precision matrix expansion



- measurement of **ξ** as a function of **x**:

$$\hat{\xi}_1 \equiv \hat{\xi}(x_1)\ ,\ \ \hat{\xi}_2 \equiv \hat{\xi}(x_2)$$

- (unknown) expectation value of the measurement:

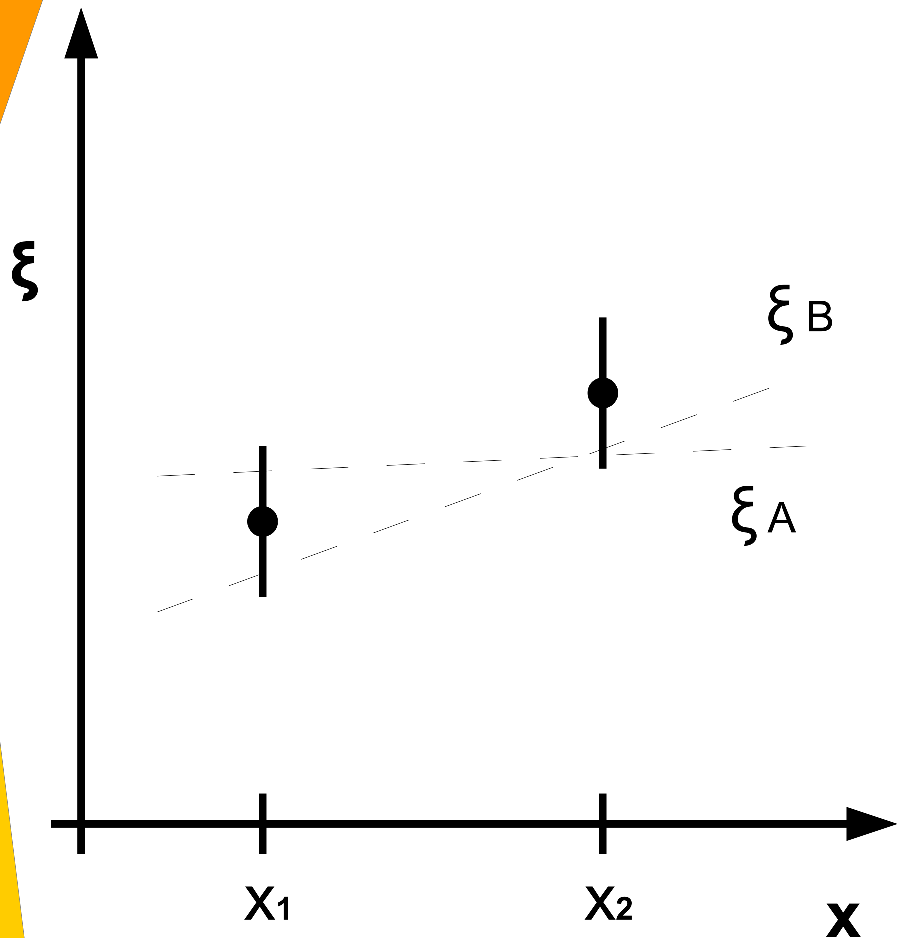
$$\xi_1 \equiv \langle\hat{\xi}_1\rangle\ ,\ \ \xi_2 \equiv \langle\hat{\xi}_2\rangle$$

- covariance matrix of the measurement:

$$\mathrm{C}_{ij} = \langle(\hat{\xi}_i - \xi_i)(\hat{\xi}_j - \xi_j)\rangle$$

$$\mathbf{C} = \begin{pmatrix} \mathrm{C}_{11} & \mathrm{C}_{12} \\ \mathrm{C}_{21} & \mathrm{C}_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

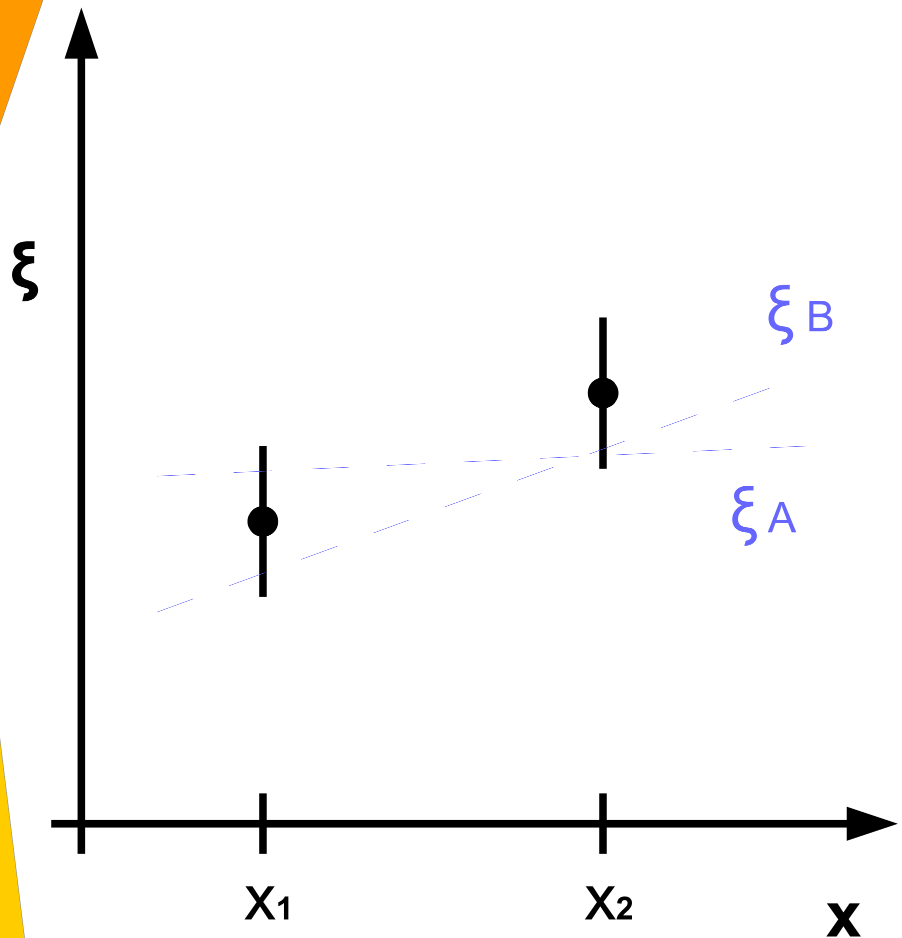
(r = Pearson correlation)



- knowledge of the complete covariance needed to judge agreement between models and data

- Which model can be ruled out? (diagonal covariance)

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

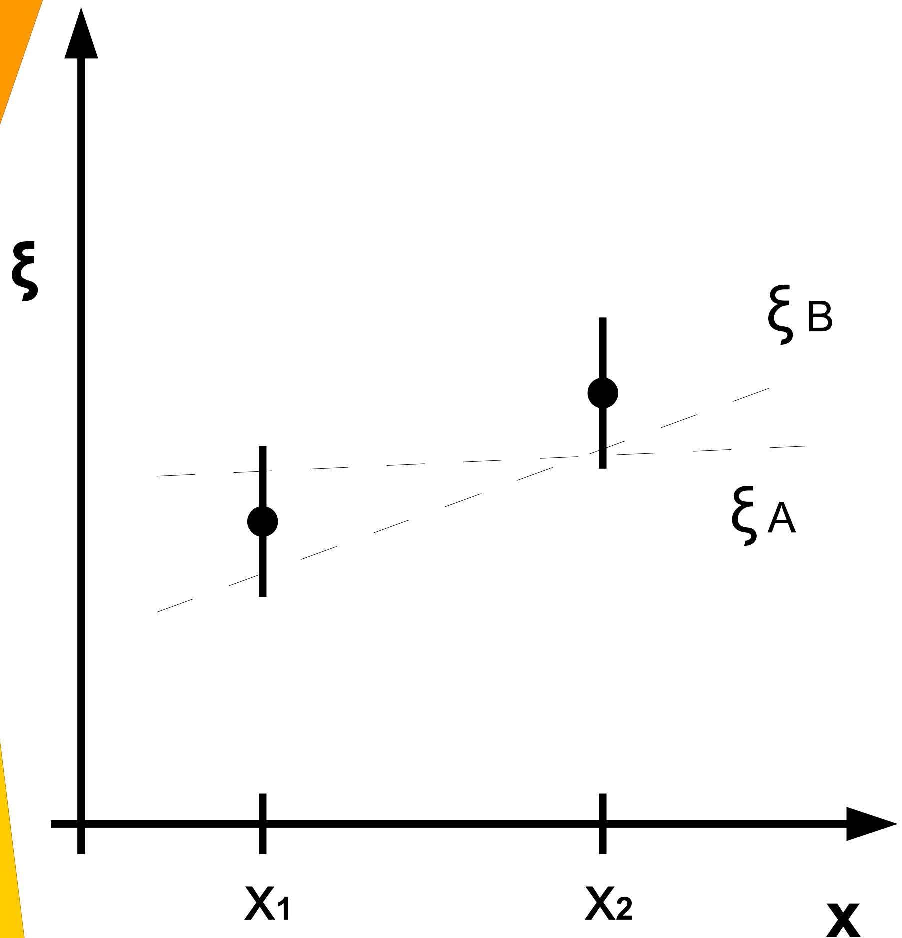


- knowledge of the complete covariance needed to judge agreement between models and data

- Which model can be ruled out? (diagonal covariance)

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

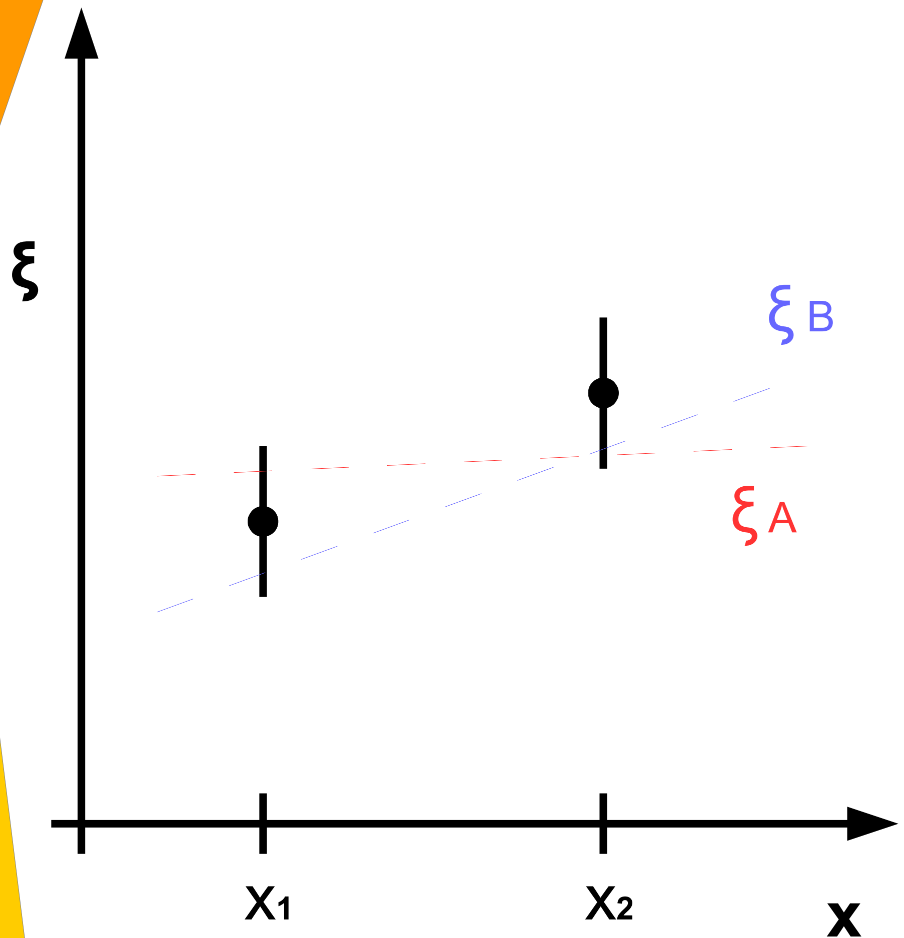
→ none



- knowledge of the complete covariance needed to judge agreement between models and data

- Which model can be ruled out? (strong positive correlation)

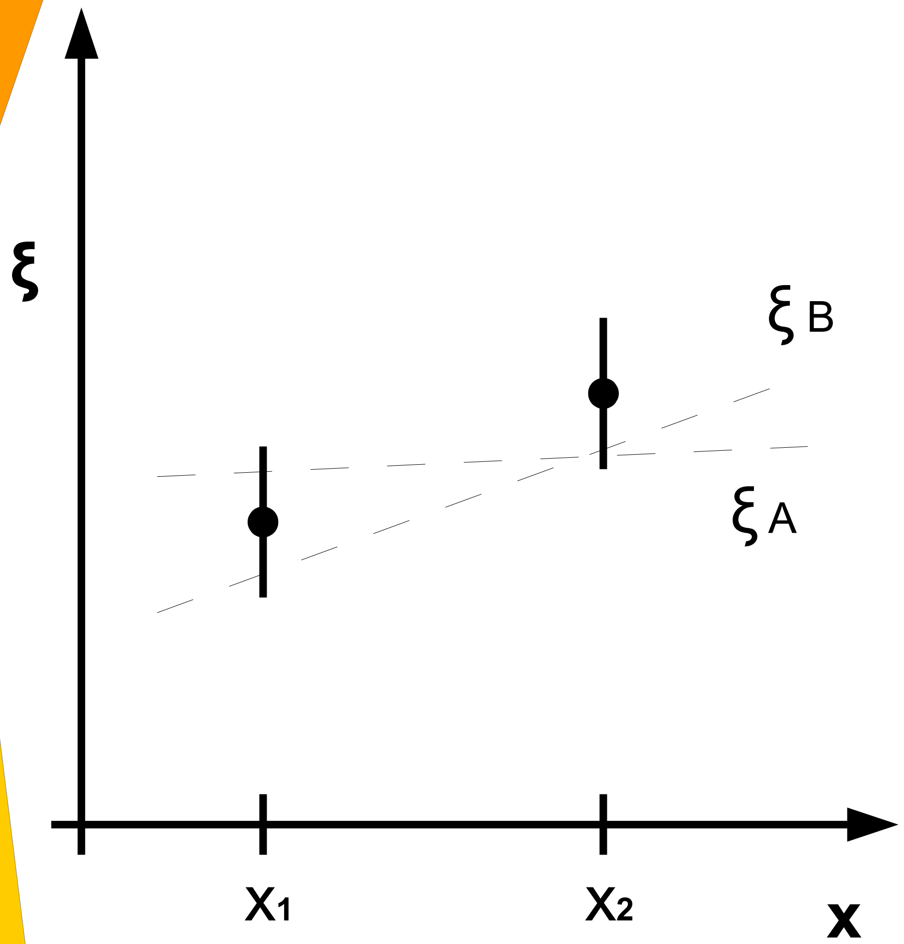
$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0.9\sigma^2 \\ 0.9\sigma^2 & \sigma^2 \end{pmatrix}$$



- knowledge of the complete covariance needed to judge agreement between models and data
- Which model can be ruled out? (strong positive correlation)

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0.9\sigma^2 \\ 0.9\sigma^2 & \sigma^2 \end{pmatrix}$$

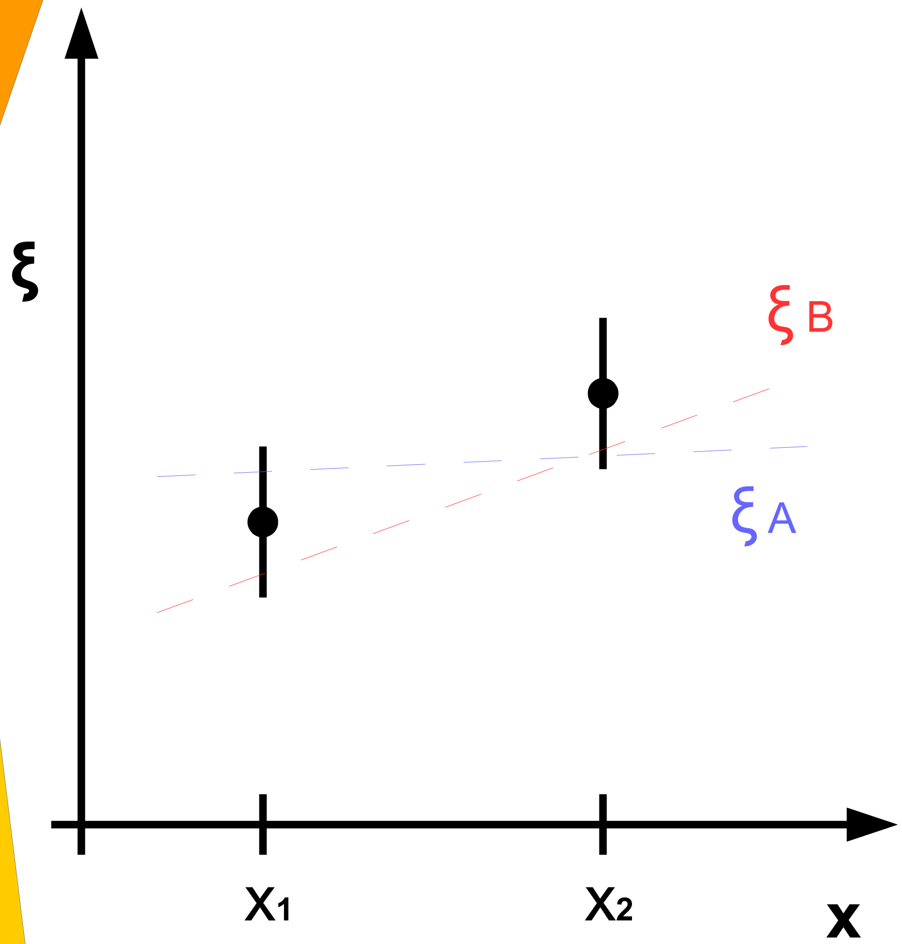
→ model A



- knowledge of the complete covariance needed to judge agreement between models and data

- Which model can be ruled out? (strong negative correlation)

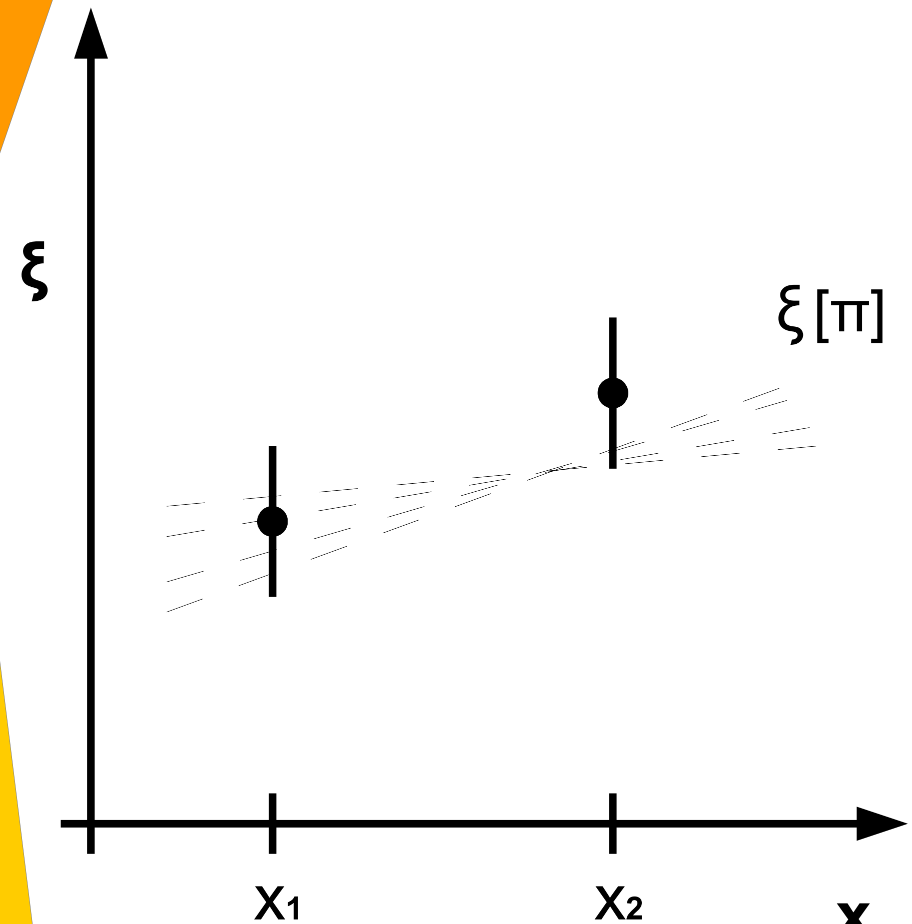
$$\mathbf{C} = \begin{pmatrix} \sigma^2 & -0.9\sigma^2 \\ -0.9\sigma^2 & \sigma^2 \end{pmatrix}$$



- knowledge of the complete covariance needed to judge agreement between models and data
- Which model can be ruled out? (strong negative correlation)

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & -0.9\sigma^2 \\ -0.9\sigma^2 & \sigma^2 \end{pmatrix}$$

→ model B



Parameter inference:

- assume that the data vector

$$\hat{\boldsymbol{\xi}} = \begin{pmatrix} \hat{\xi}_1 & \hat{\xi}_2 \end{pmatrix}$$

  has as multivariate Gaussian distribution:

$$p(\hat{\boldsymbol{\xi}}|\pi) \sim$$

$$\exp\left\{ -\frac{1}{2} \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\pi] \right)^T \mathbf{C}^{-1} \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}[\pi] \right) \right\}$$

- neglect model parameters for which the data would be unlikely, i.e. :

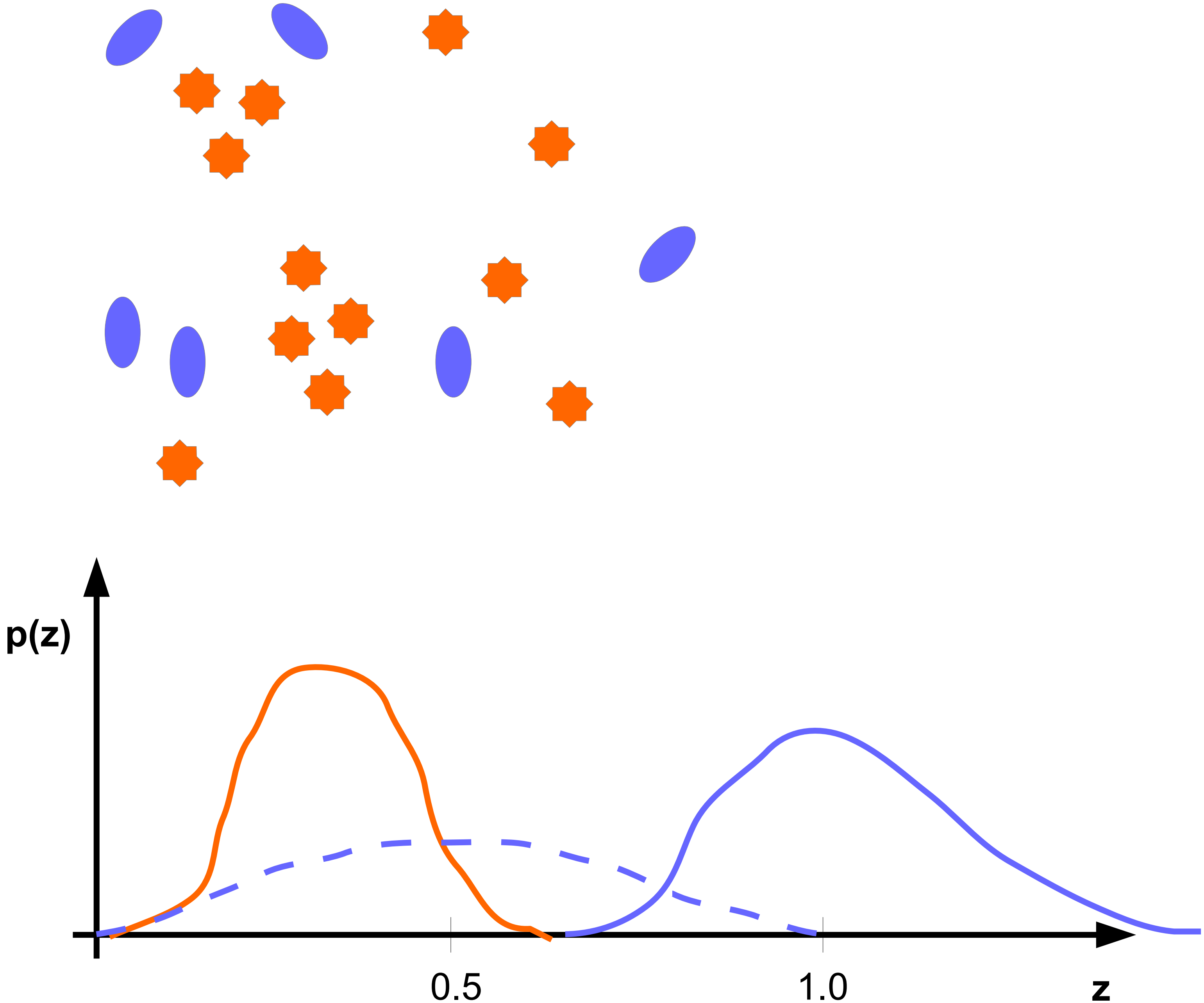
$$p(\hat{\boldsymbol{\xi}}|\pi) < \alpha$$

(Frequentist viewpoint)

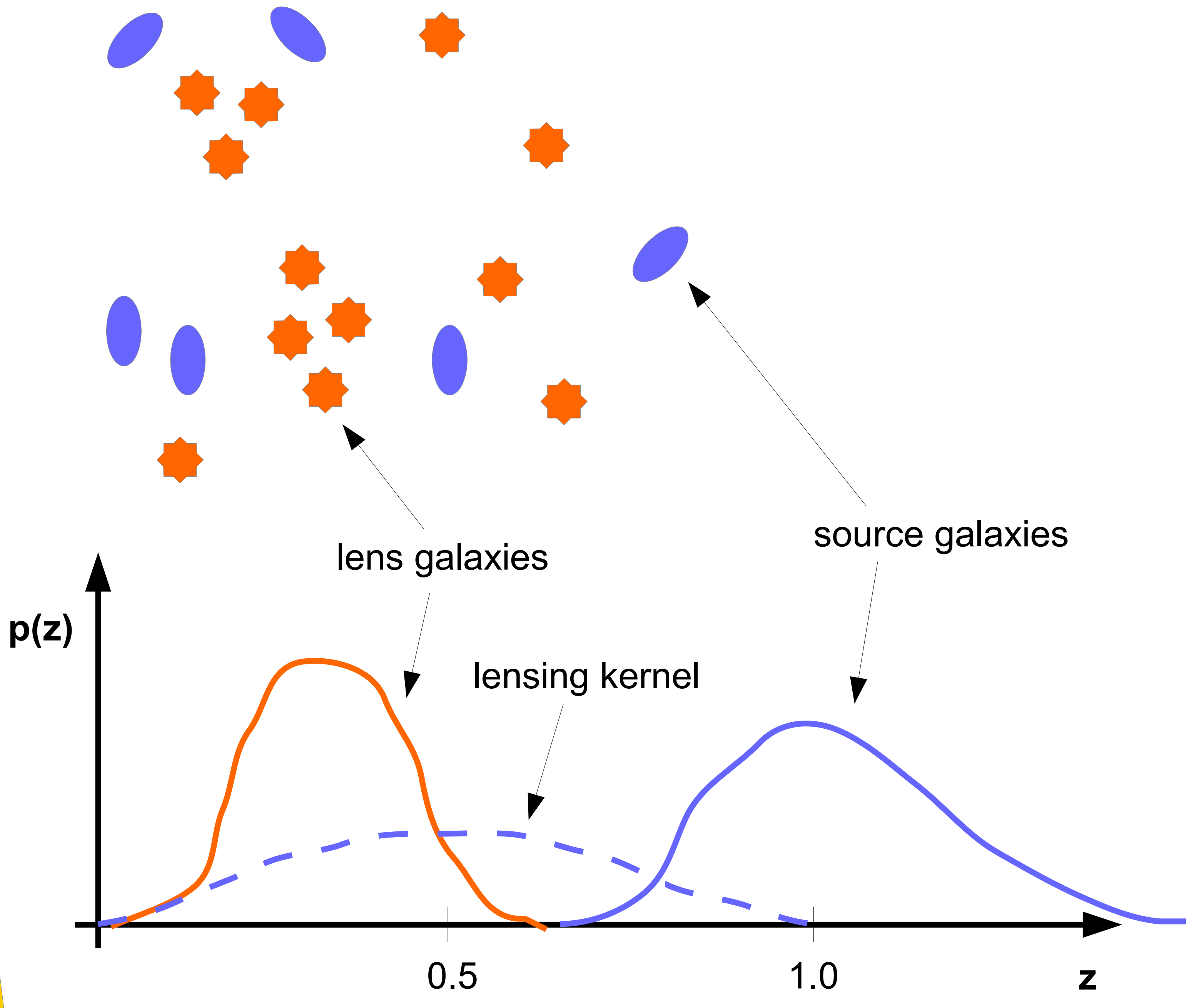
A + B = C: Estimation of the inverse covariance matrix made easy with *Precision Matrix Expansion*

- Data vectors & covariance matrices
- Multi-probe analyses
  (new orders of magnitude for covariance estimation)
- Precision matrix expansion

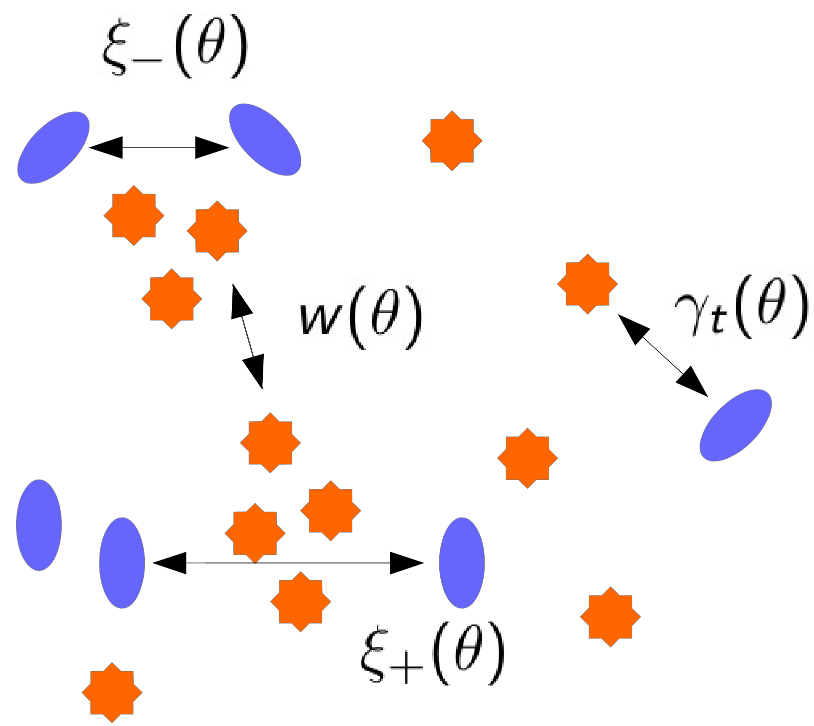
Multi-probe 2-point analyses



Multi-probe 2-point analyses

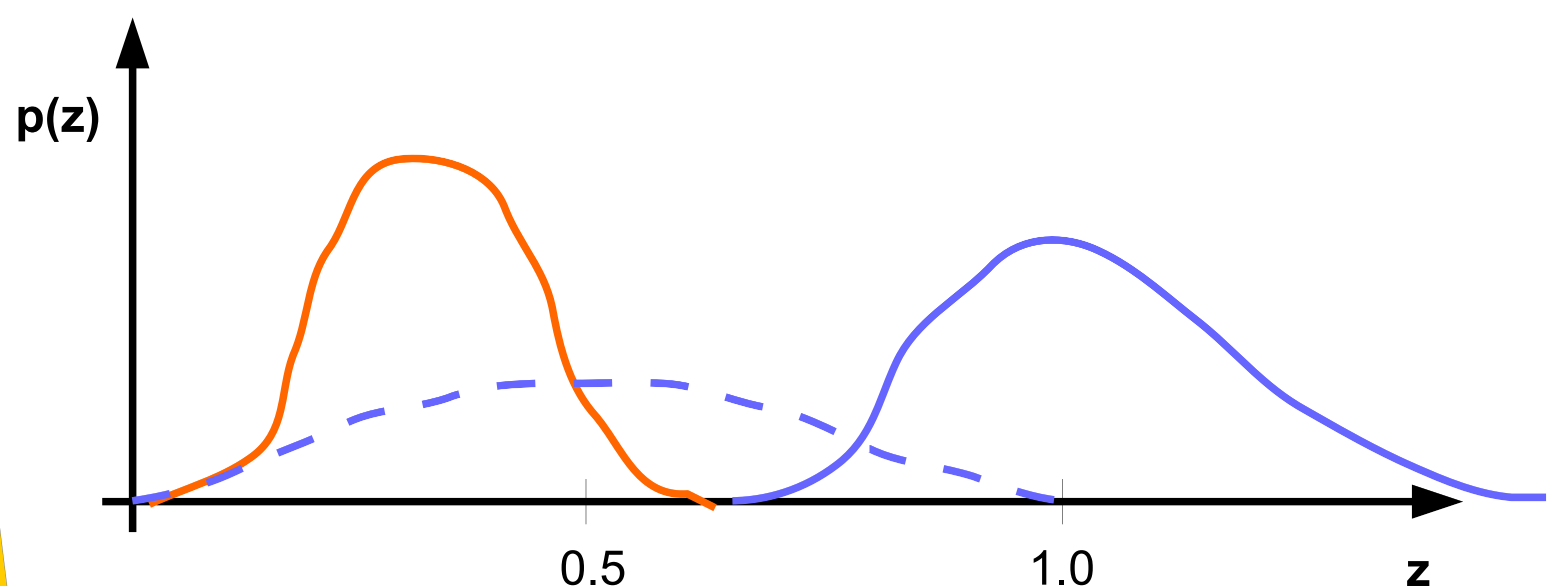


Multi-probe 2-point analyses

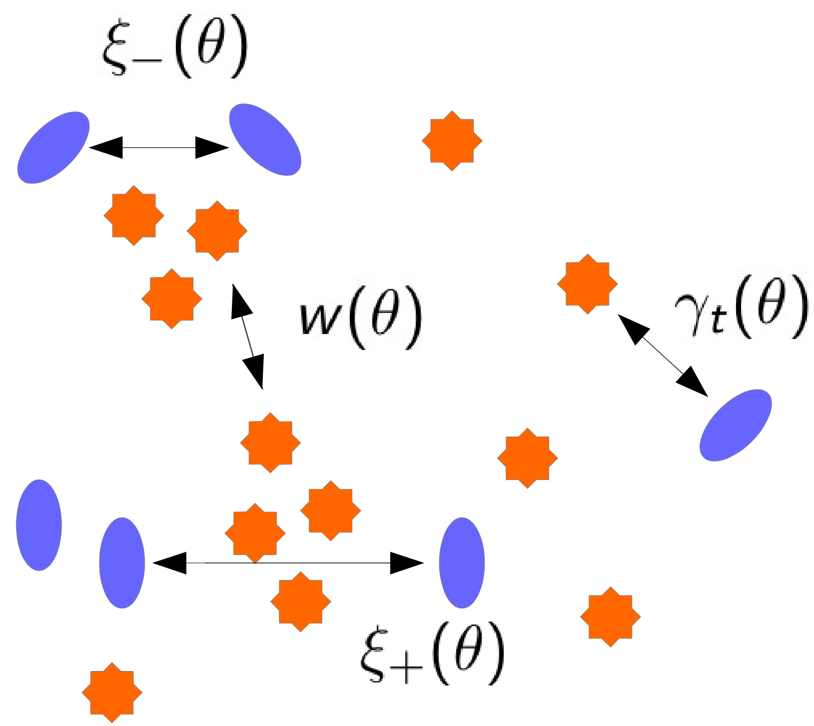


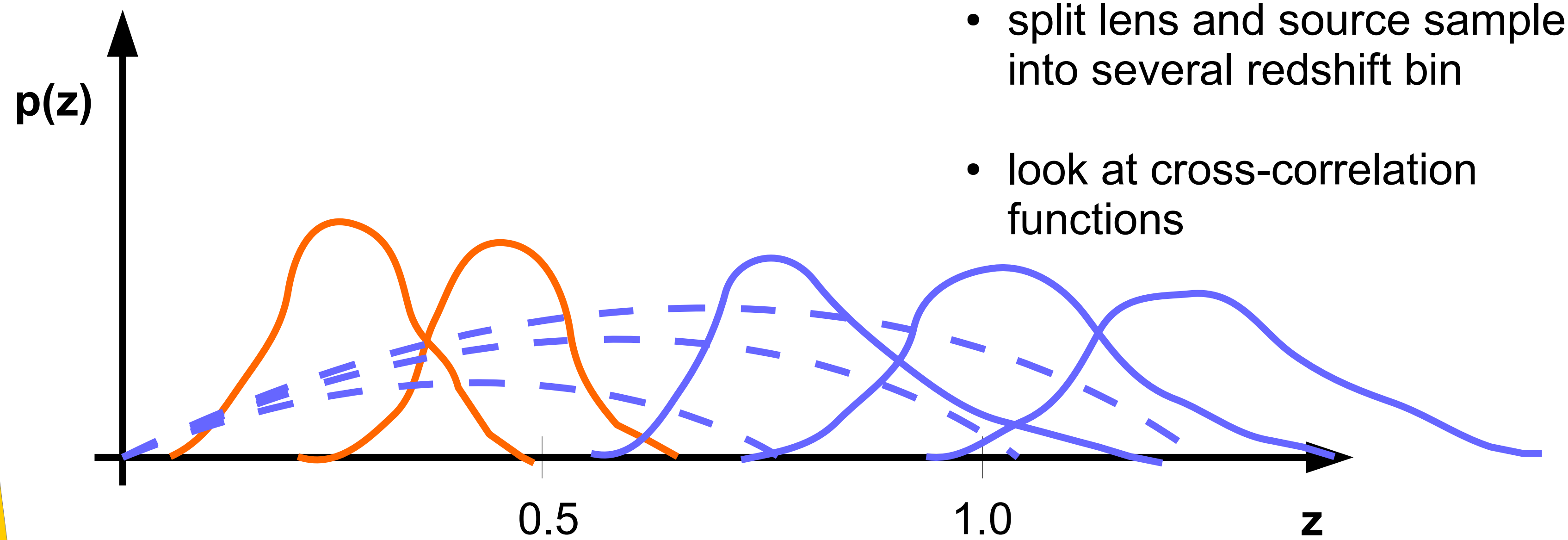
Two point correlation functions:

- galaxy clustering: $w(\theta)$
  (excess of galaxy pairs compared to a random distribution)
- galaxy-galaxy lensing: $\gamma_t(\theta)$
  (tangential stretch of source images around lens galaxies)
- cosmic shear: $\xi_+(\theta)$, $\xi_-(\theta)$
  (correlation of source galaxy shapes caused by lensing through matter inhomogeneities)



Multi-probe 2-point analyses





Two point correlation functions:

- galaxy clustering: $w(\theta)$
  (excess of galaxy pairs compared to a random distribution)
- galaxy-galaxy lensing: $\gamma_t(\theta)$
  (tangential stretch of source images around lens galaxies)
- cosmic shear: $\xi_+(\theta)$, $\xi_-(\theta)$
  (correlation of source galaxy shapes caused by lensing through matter inhomogeneities)

Redshift tomography:

- split lens and source sample into several redshift bin
- look at cross-correlation functions

In our paper:

- <u>DES weak lensing only:</u>
  450 data points

- <u>DES multi-probe:</u>
  650 data points

- <u>LSST weak lensing only:</u>
  2200 data points

→ Want to get estimates of these covariances from N-body simulations!

Why is this difficult?

We need the inverse covariance:
(the *precision matrix*)

$$\chi^2(\hat{\xi}|\mathbf{C},\pi) = \left(\hat{\xi} - \xi[\pi]\right)^T \mathbf{C}^{-1} \left(\hat{\xi} - \xi[\pi]\right)$$

In our paper:

- DES weak lensing only:
  450 data points
- DES multi-probe:
  650 data points
- LSST weak lensing only:
  2200 data points

→ Want to get estimates of these covariances from N-body simulations!

Why is this difficult?

We need the inverse covariance:
(the *precision matrix*)

$$\chi^2(\hat{\xi}|\mathbf{C},\pi) = \left(\hat{\xi} - \xi[\pi]\right)^T \mathbf{C}^{-1}\left(\hat{\xi} - \xi[\pi]\right)$$

- constructing an unbiased estimator of the precision matrix:

$$\hat{\mathrm{C}}_{ij} = \frac{1}{N_s - 1}\sum_{k=1}^{N_s}(\hat{\xi}_i^k - \bar{\xi}_i)(\hat{\xi}_j^k - \bar{\xi}_j)$$

$$\begin{aligned}\hat{\mathbf{\Psi}} &= \frac{N_s - N_d - 2}{N_s - 1}\hat{\mathbf{C}}^{-1}\\ \Rightarrow \langle\hat{\mathbf{\Psi}}\rangle &= \mathbf{C}^{-1}\end{aligned}$$

In our paper:

- <u>DES weak lensing only:</u>
  450 data points
- <u>DES multi-probe:</u>
  650 data points
- <u>LSST weak lensing only:</u>
  2200 data points

→ Want to get estimates of these covariances from N-body simulations!

Why is this difficult?

We need the inverse covariance:
(the *precision matrix*)

$$\chi^2(\hat{\xi}|\mathbf{C},\pi) = \left(\hat{\xi} - \xi[\pi]\right)^T \mathbf{C}^{-1}\left(\hat{\xi} - \xi[\pi]\right)$$

- constructing an unbiased estimator of the precision matrix:

$$\hat{C}_{ij} = \frac{1}{N_s - 1}\sum_{k=1}^{N_s}(\hat{\xi}_i^k - \bar{\xi}_i)(\hat{\xi}_j^k - \bar{\xi}_j)$$

$$\hat{\mathbf{\Psi}} = \frac{N_s - N_d - 2}{N_s - 1}\hat{\mathbf{C}}^{-1}$$

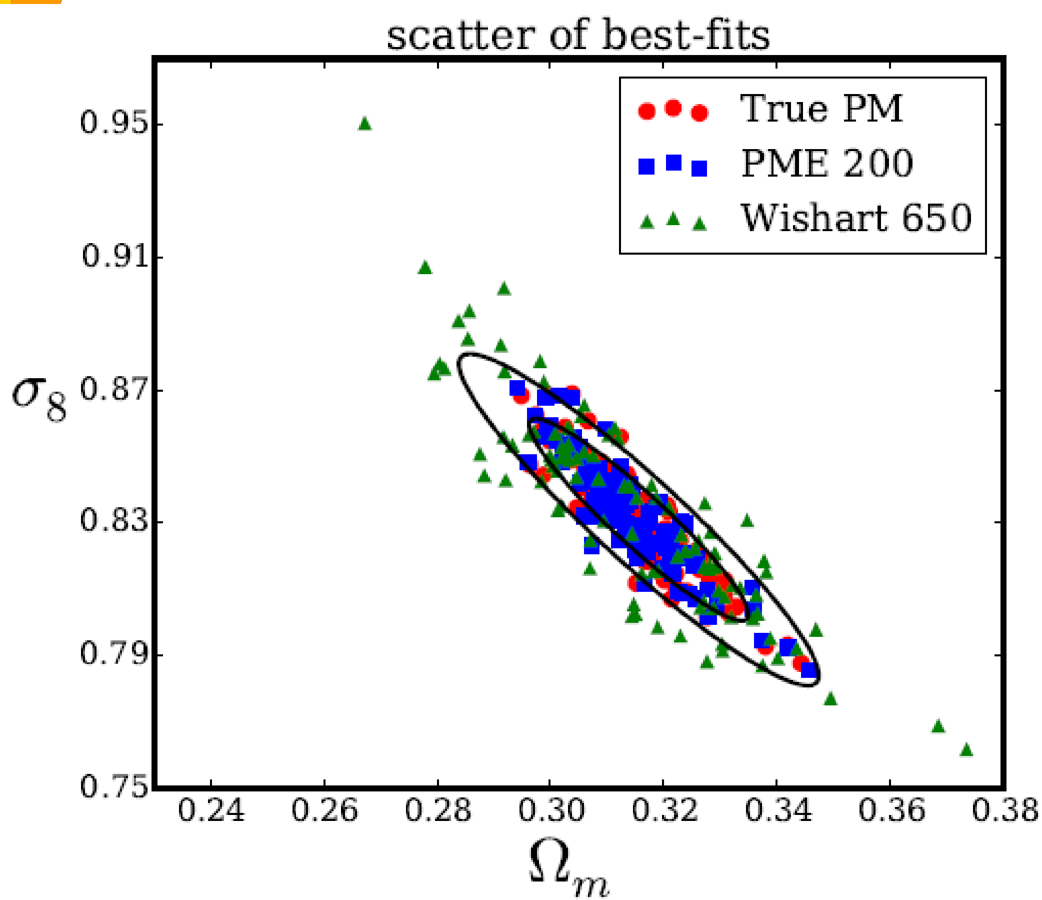
$$\Rightarrow \langle\hat{\mathbf{\Psi}}\rangle = \mathbf{C}^{-1}$$

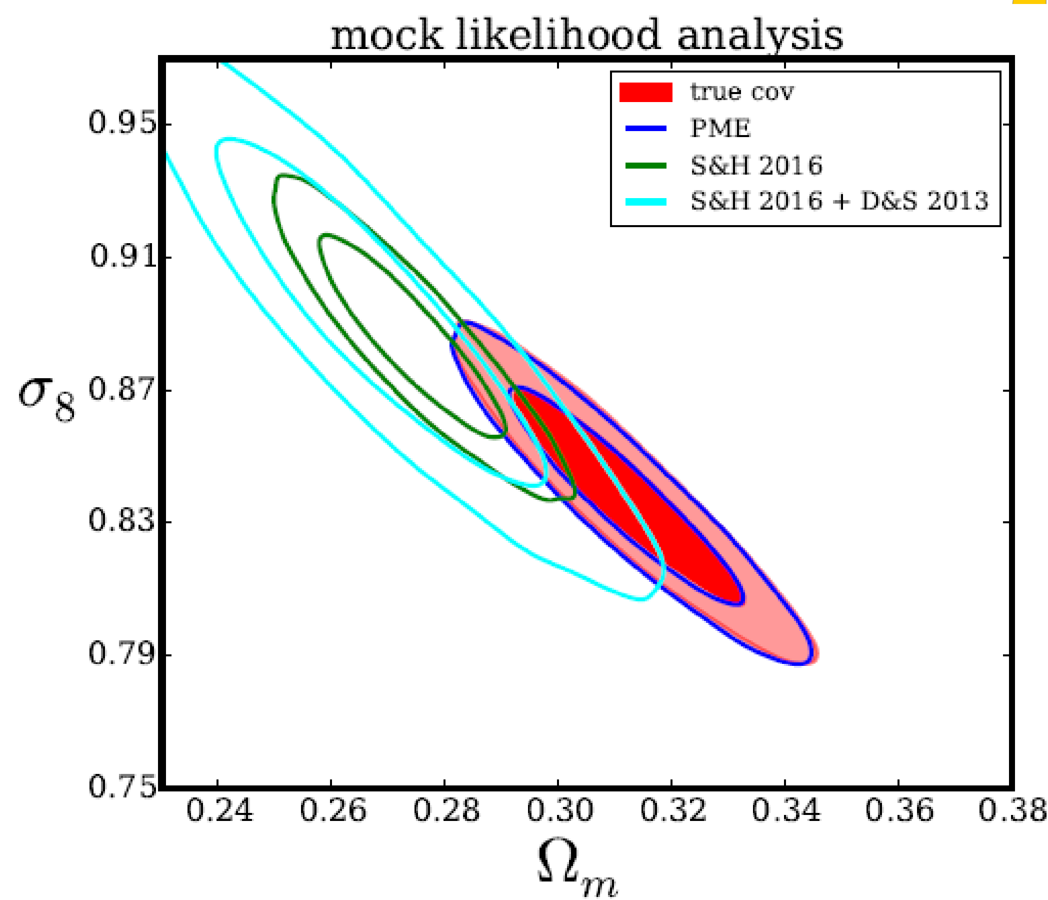
- estimation noisy of covariance and precision matrix:

$$\frac{\Delta\hat{C}_{ii}}{\hat{C}_{ii}} \approx \sqrt{\frac{2}{N_s - 1}}$$

$$\frac{\Delta\hat{\Psi}_{ii}}{\hat{\Psi}_{ii}} \approx \sqrt{\frac{2}{N_s - N_d - 2}}$$

View parameter inference as a 2-step process:



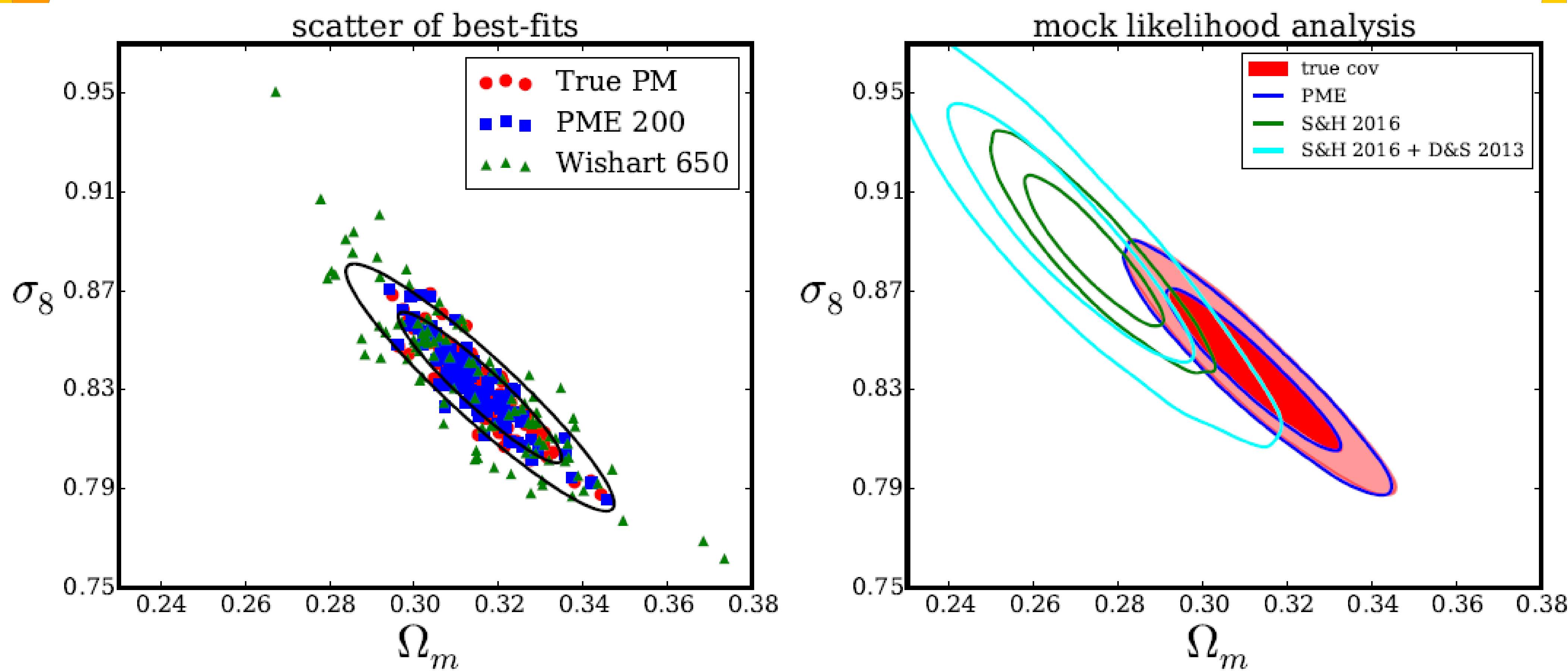


1. determine the best-fit parameters

2. draw confidence regions around them

→ both steps use $$\chi^2(\hat{\xi}|\mathbf{C},\pi) = \left(\hat{\xi} - \xi[\pi]\right)^T \mathbf{C}^{-1} \left(\hat{\xi} - \xi[\pi]\right)$$

View parameter inference as a 2-step process:



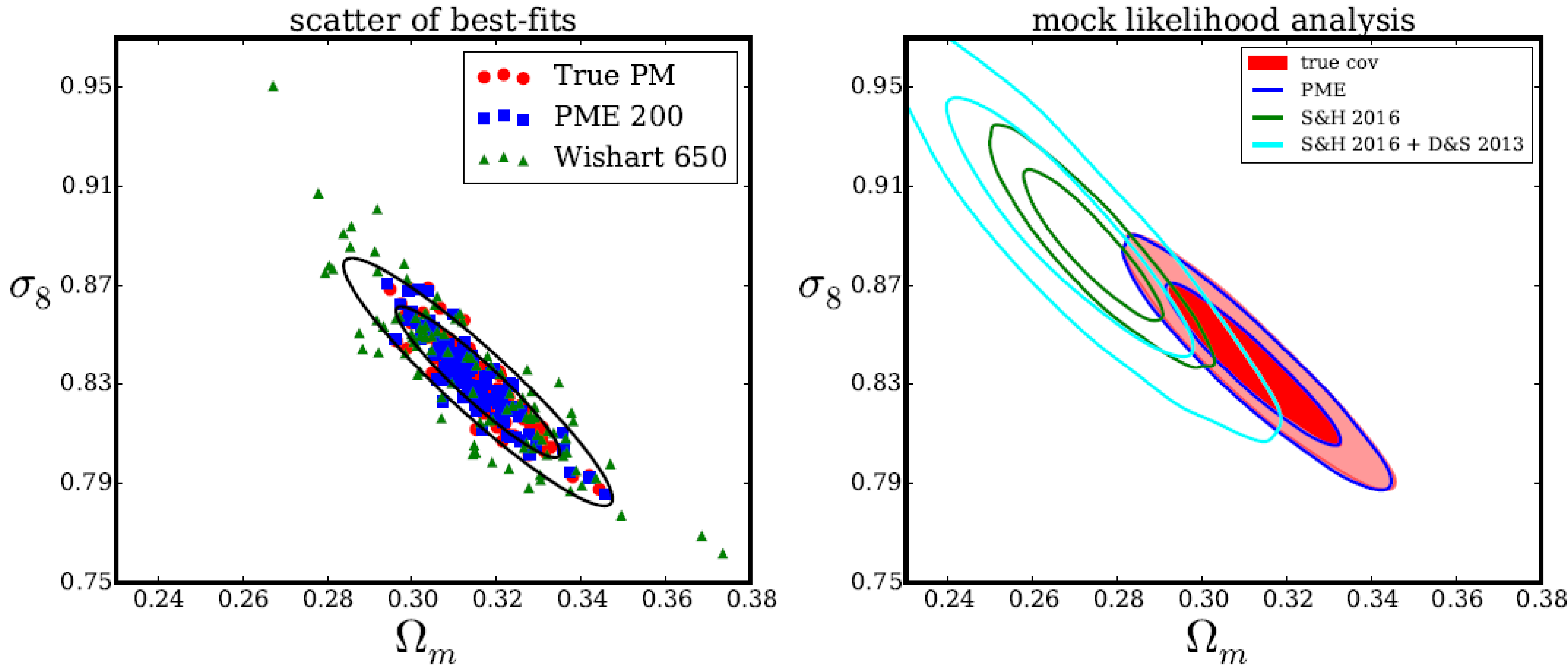
1. determine the best-fit parameters

2. draw confidence regions around them

→ both steps use

$$\chi^2(\hat{\xi}|\hat{\boldsymbol{\Psi}}, \pi) = \left(\hat{\xi} - \xi[\pi]\right)^T \hat{\boldsymbol{\Psi}} \cdot \left(\hat{\xi} - \xi[\pi]\right)$$

View parameter inference as a 2-step process:



1. determine the best-fit parameters

$$\frac{\tilde{\Delta}\pi}{\sigma_\pi} \sim \sqrt{\frac{N_d - N_p}{N_s - N_d}}$$

2. draw confidence regions around them

$$\frac{\Delta\sigma_\pi^2}{\sigma_\pi^2} \sim \sqrt{\frac{1}{N_s - N_d}}$$

A + B = C: Estimation of the inverse covariance matrix made easy with *Precision Matrix Expansion*

- Data vectors & covariance matrices
- Multi-probe analyses
  (new orders of magnitude for covariance estimation)
- Precision matrix expansion

$$\frac{1}{1+x} = 1 - x + x^2 - \mathcal{O}(x^3)$$

1.) Avoid matrix inversion with a priori knowledge about the covariance

- true covariance: $\mathbf{C}$
- covariance estimate: $\hat{\mathbf{C}}$
- model covariance: $\mathbf{M}$
- 'relative' deviation between model and true covariance:

$$\mathbf{X} = (\mathbf{C} - \mathbf{M})\ \mathbf{M}^{-1}$$

$$\hat{\mathbf{X}} = (\hat{\mathbf{C}} - \mathbf{M})\ \mathbf{M}^{-1}$$

Precision matrix expansion:

$$\begin{aligned} \mathbf{C} &= \mathbf{M} + \mathbf{C} - \mathbf{M} \\ &= (\mathbf{I} + \mathbf{X})\ \mathbf{M} \\ \Rightarrow \mathbf{\Psi} &= \mathbf{M}^{-1}\ [\mathbf{I} - \mathbf{X} + \mathbf{X}^2 - \mathcal{O}(\mathbf{X}^3)] \end{aligned}$$

- estimation of the first order term:

$$\hat{\mathbf{\Psi}}_{1st} = 2\mathbf{M}^{-1} - \mathbf{M}^{-1}\hat{\mathbf{C}}\ \mathbf{M}^{-1}$$

has noise $\sim \sqrt{\dfrac{2}{N_s - 1}}$

→ much less noisy than standard estimator
(but bias due to finite break of the series)

2.) Use simulations only for the covariance parts where you really need them

- constributions like shape-noise can accurately be modelled analytically!
- Split covariance as

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

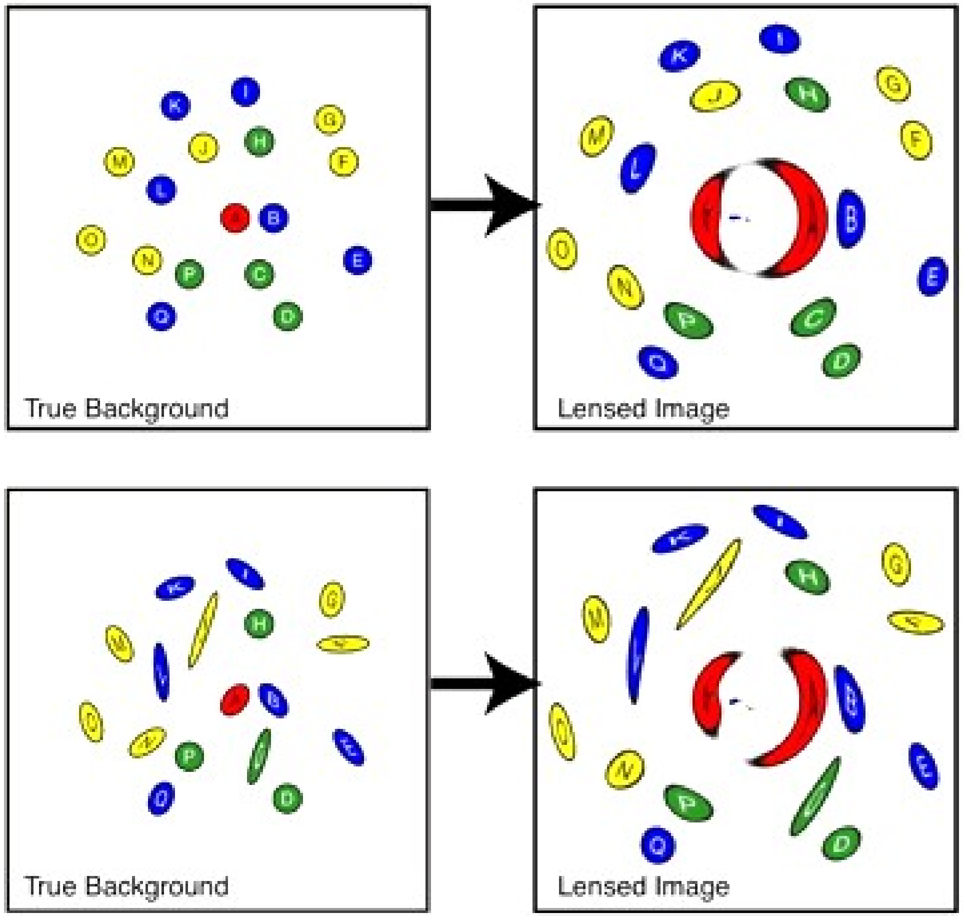
where

$\mathbf{A}$ can be modelled accurately

and can be turned off in simulations

- let $\mathbf{B}_m$ be a model for $\mathbf{B}$ and

$\mathbf{M} = \mathbf{A} + \mathbf{B}_m$ the total model



www.jyi.org/

e.g.: $\mathbf{A}$ = shape-noise in weak lensing

2.) Use simulations only for the covariance parts where you really need them

- constributions like shape-noise can accurately be modelled analytically!
- Split covariance as

$$\mathsf{C} = \mathsf{A} + \mathsf{B}$$

where

A can be modelled accurately

and can be turned off in simulations

- let B_m be a model for B and

$\mathsf{M} = \mathsf{A} + \mathsf{B}_m$ the total model

- Precision matrix expansion:

$$\Rightarrow \boldsymbol{\Psi} = \mathsf{M}^{-1} \left[\mathsf{I} - \mathsf{X} + \mathsf{X}^2 - \mathcal{O}(\mathsf{X}^3)\right]$$

with

- $$\mathsf{X} = (\mathsf{B} - \mathsf{B}_m)\, \mathsf{M}^{-1}$$

- estimation of the first order term:

$$\hat{\boldsymbol{\Psi}}_{1\mathrm{st}} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1}$$

$\rightarrow$ noisy part of the estimator becomes even smaller

2.) Use simulations only for the covariance parts where you really need them

- constributions like shape-noise can accurately be modelled analytically!
- Split covariance as

$$\mathsf{C} = \mathsf{A} + \mathsf{B}$$

where

A can be modelled accurately

and can be turned off in simulations

- let B_m be a model for B and

$\mathsf{M} = \mathsf{A} + \mathsf{B}_m$ the total model

- Precision matrix expansion:

$$\Rightarrow \boldsymbol{\Psi} = \mathsf{M}^{-1}\left[\mathsf{I} - \mathsf{X} + \mathsf{X}^2 - \mathcal{O}(\mathsf{X}^3)\right]$$

with

- $$\mathsf{X} = (\mathsf{B} - \mathsf{B}_m)\,\mathsf{M}^{-1}$$

- estimation of the second order term: (with help of Letac and Massam 2004)

$$\begin{aligned}
\hat{\Psi}_{\rm 2nd} = \; & \mathbf{M}^{-1} + \mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\
& -\mathbf{M}^{-1}\left(\hat{\mathbf{B}} - \mathbf{B}_m\right)\mathbf{M}^{-1} \\
& -\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1} \\
& -\mathbf{M}^{-1}\mathbf{B}_m\mathbf{M}^{-1}\hat{\mathbf{B}}\mathbf{M}^{-1} \\
& +\mathbf{M}^{-1}\frac{\nu^2\hat{\mathbf{B}}\mathbf{M}^{-1}\hat{\mathbf{B}} - \nu\hat{\mathbf{B}}\,\mathrm{tr}\left(\mathbf{M}^{-1}\hat{\mathbf{B}}\right)}{\nu^2 + \nu - 2}\mathbf{M}^{-}
\end{aligned}$$

Results and open questions

- Even strong deviations of our fiducial covariance seems to yield a convergent series
- For DES cosmic shear the PME needs only 200 simulations to perform as the standard estimator with > 8000 sims
- slightly worse performance for DES multi-probe
- For LSST cosmic shear the PME needs only 2200 simulations to perform as the standard estimator with > 115.000 sims

- How do we know a priori, whether the series converges?
- Can we make use of other noise terms such as shot-noise?
- How do we account for the remaining additional scatter in best-fit parameters?

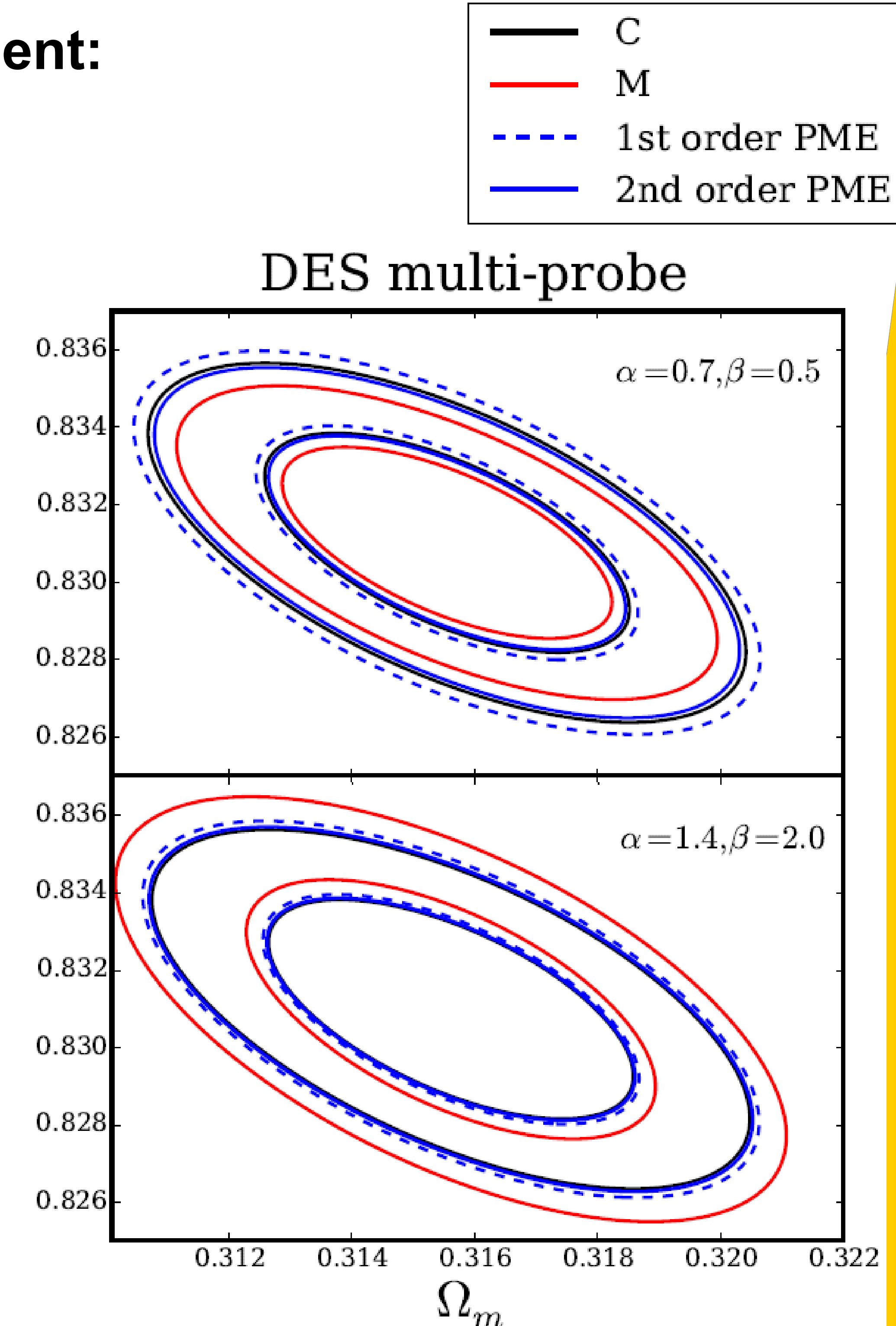
Show paper!

Results from mock experiment:

- Use halo model covariance by Krause & Eifler (2016) as **C**
- Deform it in different ways to produce a fake model covariance **M**

  (through rescaling of some covariance parts & more complicated procedures)

Does the PME converge??

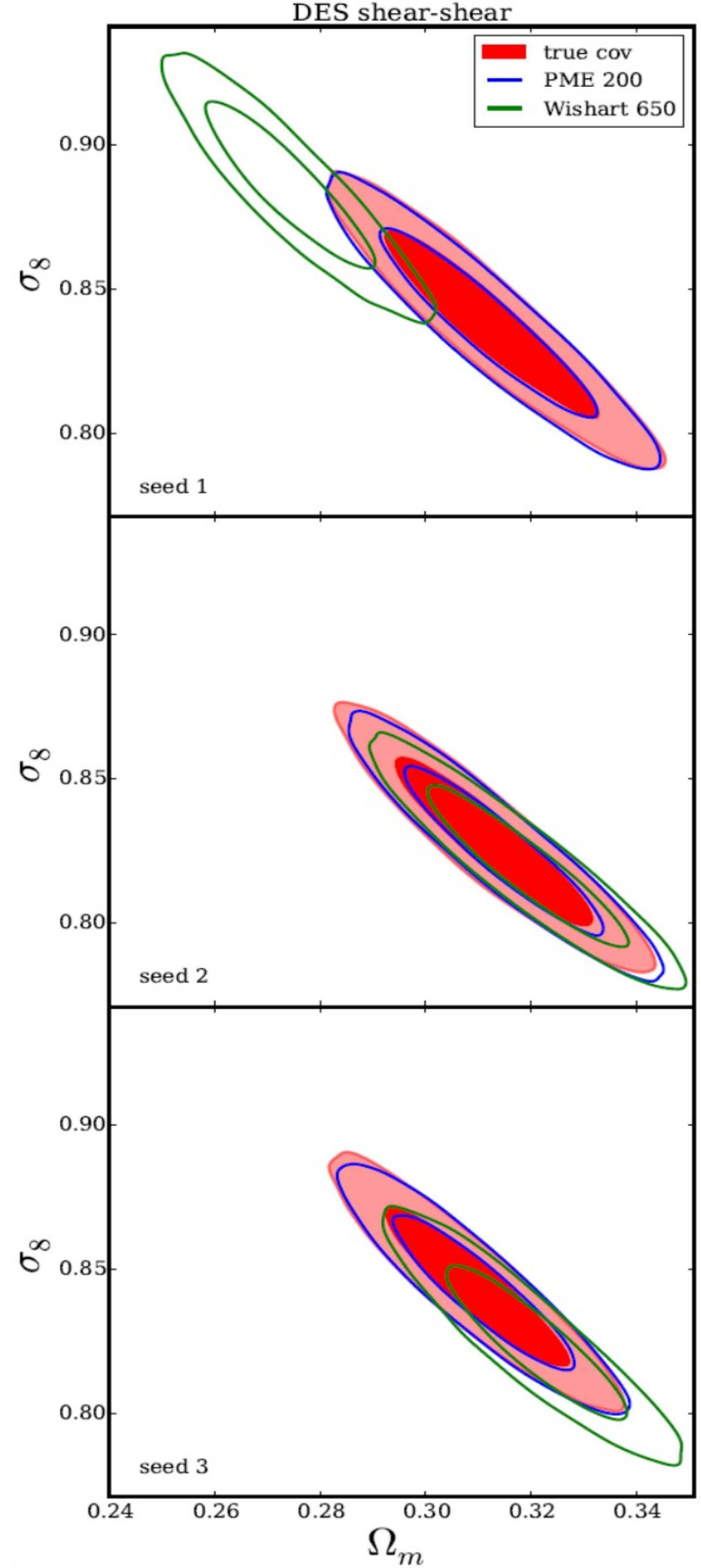


Results from mock experiment:

- Use halo model covariance by Krause & Eifler (2016) as $\mathbf{C}$
- Deform it in different ways to produce a fake model covariance $\mathbf{M}$

  (through rescaling of some covariance parts & more complicated procedures)

How does it perform with finite number of sims??



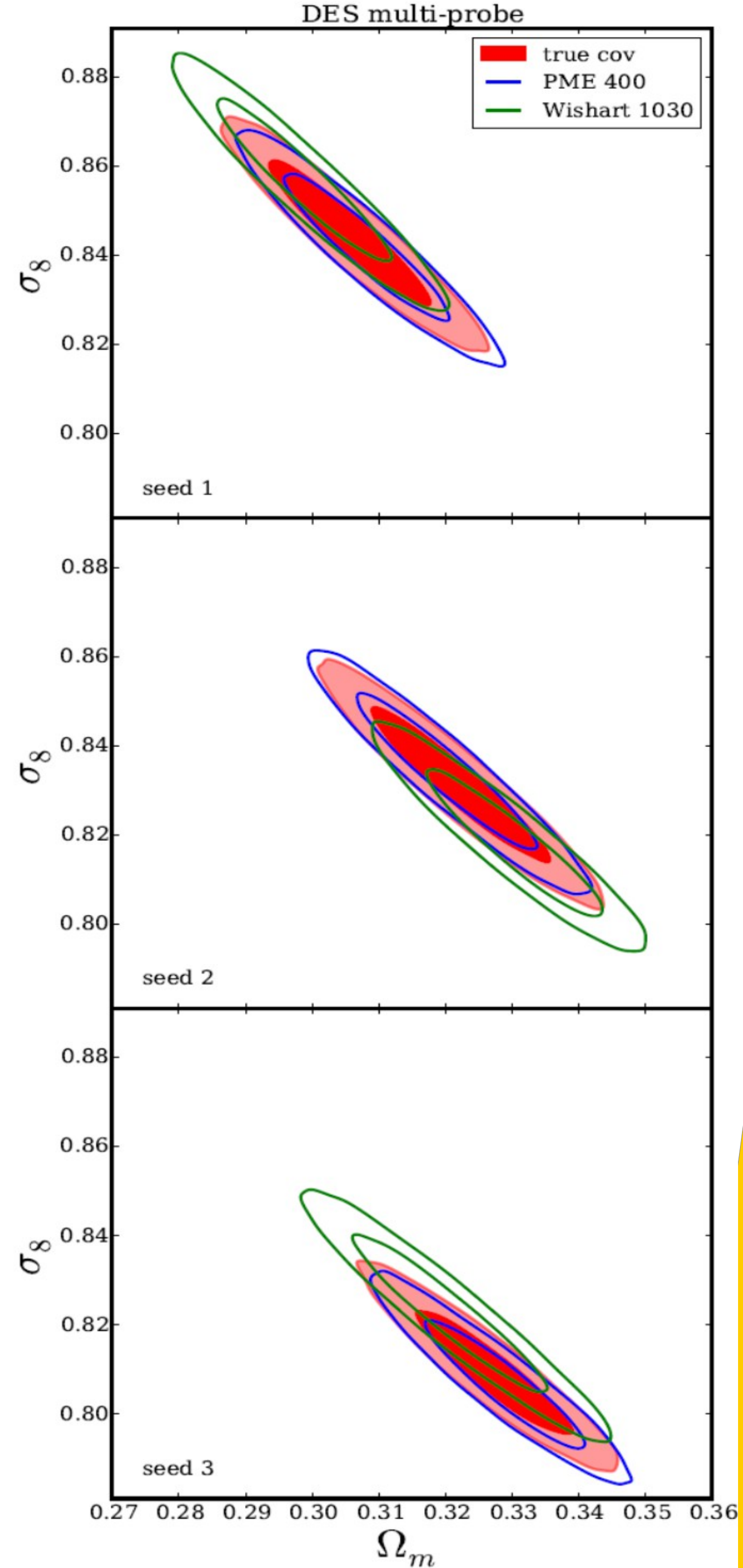
Results from mock experiment:

- Use halo model covariance by Krause & Eifler (2016) as
  C

- Deform it in different ways to produce a fake model covariance
  M

  (through rescaling of some covariance parts & more complicated procedures)

How does it perform with finite number of sims??



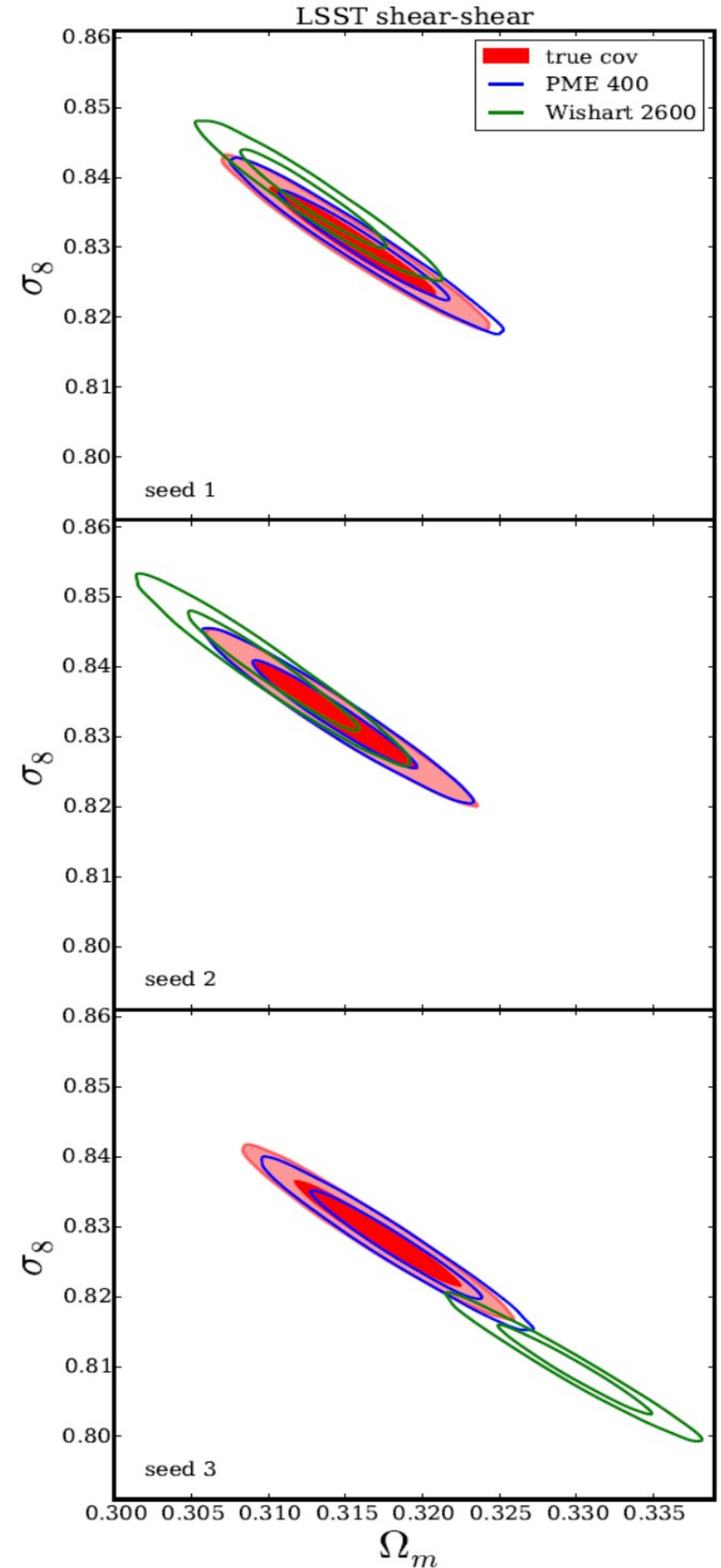
Results from mock experiment:

- Use halo model covariance by Krause & Eifler (2016) as
  $\mathbf{C}$

- Deform it in different ways to produce a fake model covariance
  $\mathbf{M}$

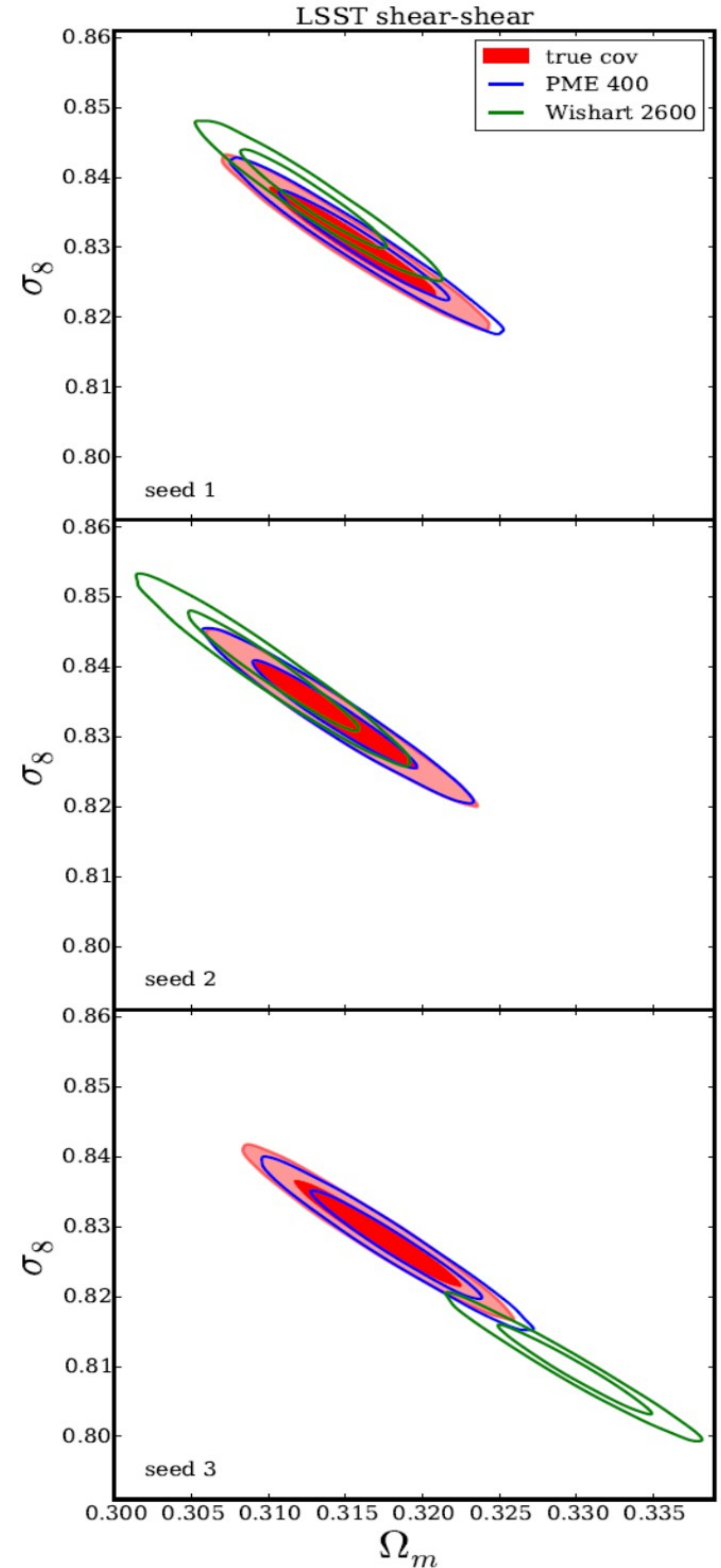
  (through rescaling of some covariance parts & more complicated procedures)

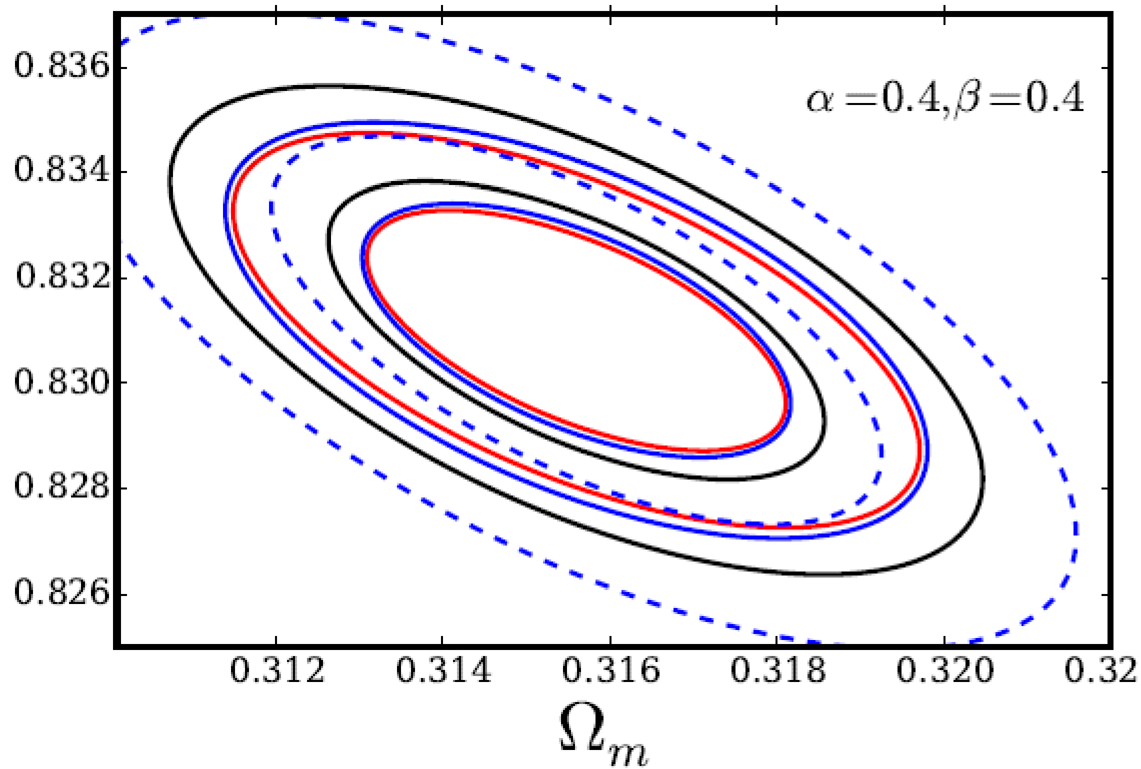
How does it perform with finite number of sims??



Conclusions

- PME robust towards strong deviations between model and N-body covariance
- for weak lensing only: excellent recovery of parameter constraints
- for galaxy clustering:
  still big improvement as opposed to standard way to estimate precision matrix





What if the PME does not converge??