Why Public Likelihoods?
The EFT viewpoint

Juan Rojo
VU Amsterdam & Theory group, Nikhef

Towards a white paper on public likelihoods -3-

12/05/2021

most issues common with those of PDF fits - see Robert's talk!

Global EFT interpretations

consider general expression of a LHC cross-section in the **SMEFT**

$$\sigma_{\rm LHC}\left(\boldsymbol{c}, \Lambda, \boldsymbol{\theta}\right) \simeq \left(\int_{M^2}^{s} d\hat{s}\; \mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})\, \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s) \right) \times \left(1 + \sum_{m=1}^{N_6} c_m \frac{\kappa_m}{\Lambda^2} + \sum_{m,n=1}^{N_6} c_m c_n \frac{\kappa_{mn}}{\Lambda^4} \right) ,$$

PDF parameters ***EFT coefficients***

- Set to zero EFT coefficients, fit PDF parameters: **global PDF analysis** ***Robert's talk!***
- Fix PDFs from some datasets, use LHC data to constrain EFT coefficients: **global EFT fit** ***this talk!***
- **Simultaneous determination** of PDFs with EFT coefficients ***Greljo et al. 21***

What is most relevant, in this context, about the global EFT fitting program?
in addition to issues raised by Robert

- ☑ One aims to coherently interpret a **wide range of measurements** from the LHC
  common statistical model for all measurements crucial
- ☑ Use results to identify **optimally sensitive measurements**, including at detector-level
  fully analog to searches recasting

Global EFT interpretations

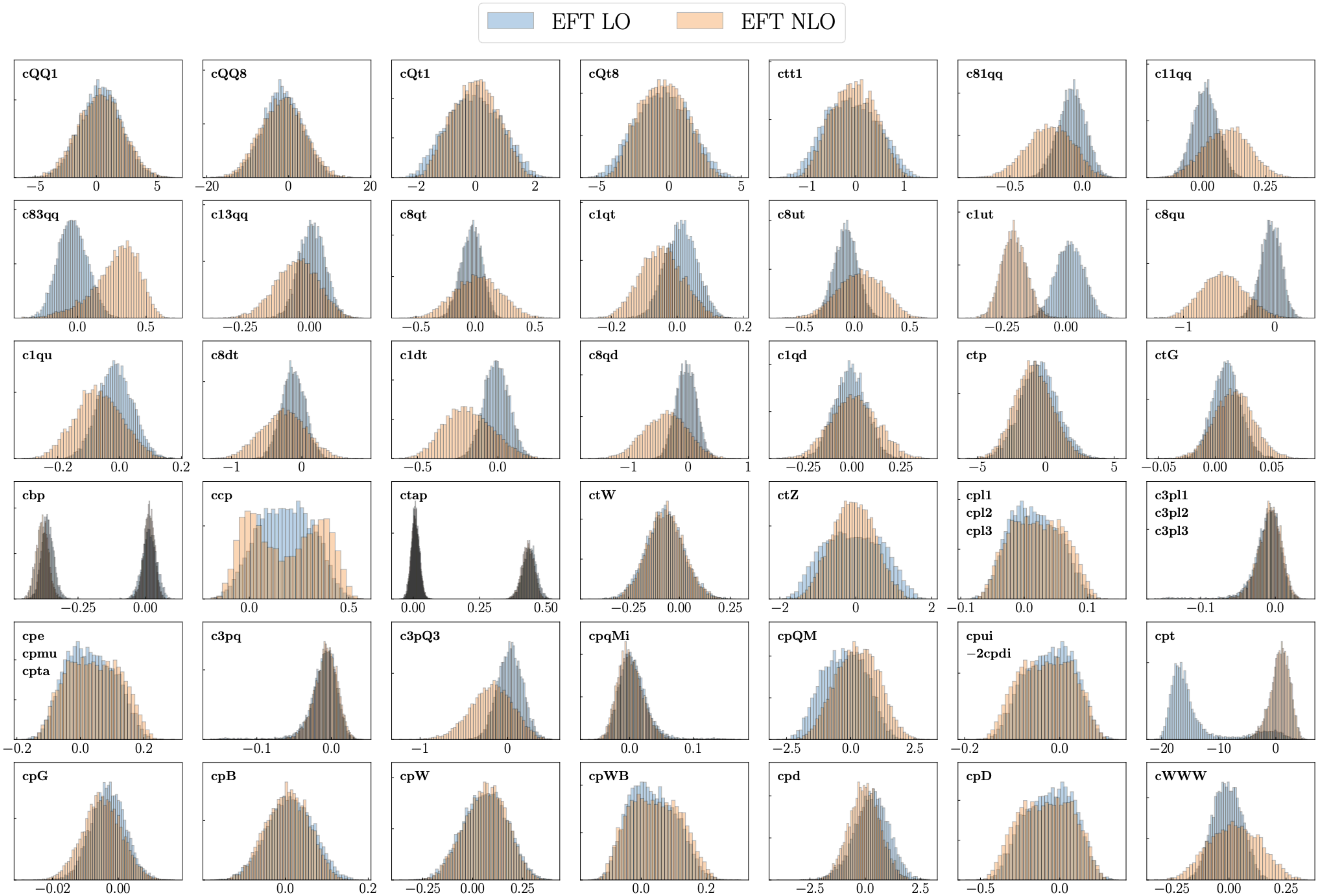
- Simplest option: determine Wilson coefficients from **log-likelihood minimisation**

$$\chi^2(\{c_n^{(k)}\}) \equiv \frac{1}{N_{\rm dat}} \sum_{i,j=1}^{N_{\rm dat}} \left(\mathcal{O}_i^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_i^{(\rm exp)} \right) ({\rm cov}^{-1})_{ij} \left(\mathcal{O}_j^{(\rm th)}\left(\{c_n^{(k)}\}\right) - \mathcal{O}_j^{(\rm exp)} \right)$$

- Assumes that all uncertainties (experimental and theory) are **Gaussianly distributed**
- Even at this level, a global EFT analysis is often **limited** by:
  - Lack of information on correlations
  - Non-positive-definite covariance matrices
  - Lack of breakdown of correlated systematic sources
  - Presence of systematic sources (e.g. modelling) which might not be Gaussian
  - Different naming for systematic sources, which complicates combining processes
  - Data not available from HepData, multiple iterations with conveners necessary
  - …..

all these issues would be solved once and for all if likelihoods/statistical models were **public**

Global EFT interpretations



currently hidden under the carpet, what **impact** they have on existing and future EFT fits?

Global EFT interpretations

- Next-to-simplest option: determine Wilson coefficients from the maximisation of a **non-gaussian likelihood**

binned likelihoods, using our "convention"

$$\mu_i = \bar{\mu}_i \pm \sigma_{\text{poiss,i}} \pm \sigma_{\text{syst,gauss,i}} \pm \sum \sigma_{\text{syst,gauss,ij}} \pm \sigma_{\text{theo,ij}}$$

central value — *stat error (poisson)* — *syst error (uncorrelated, gaussian)* — *syst error (correlated, gaussian)* — *theory error (gaussian? flat?)*

- Advantage: treat all sources of theory and experimental errors with their correct distributions

$$\max_{c_n} \mathscr{L}\left(\mu_i, c_n\right) = \mathscr{L}_{\text{gauss}} \times \mathscr{L}_{\text{poiss}} \times \mathscr{L}_{\text{flat}} \times \ldots$$

- Disadvantage: experimental information seldom provided in this **format,** e.g. in **search data** one often only finds the number of events but not the breakdown of systematics

there is never **nothing lost** if information is provided this way! e.g. one can always approximate errors as Gaussian and reconstruct the covariance matrix required in the previous methods based on **χ^2 minimisation**

Global EFT interpretations

- Next-to-next-to simplest option: extract Wilson coefficients directly at the **detector-level events**

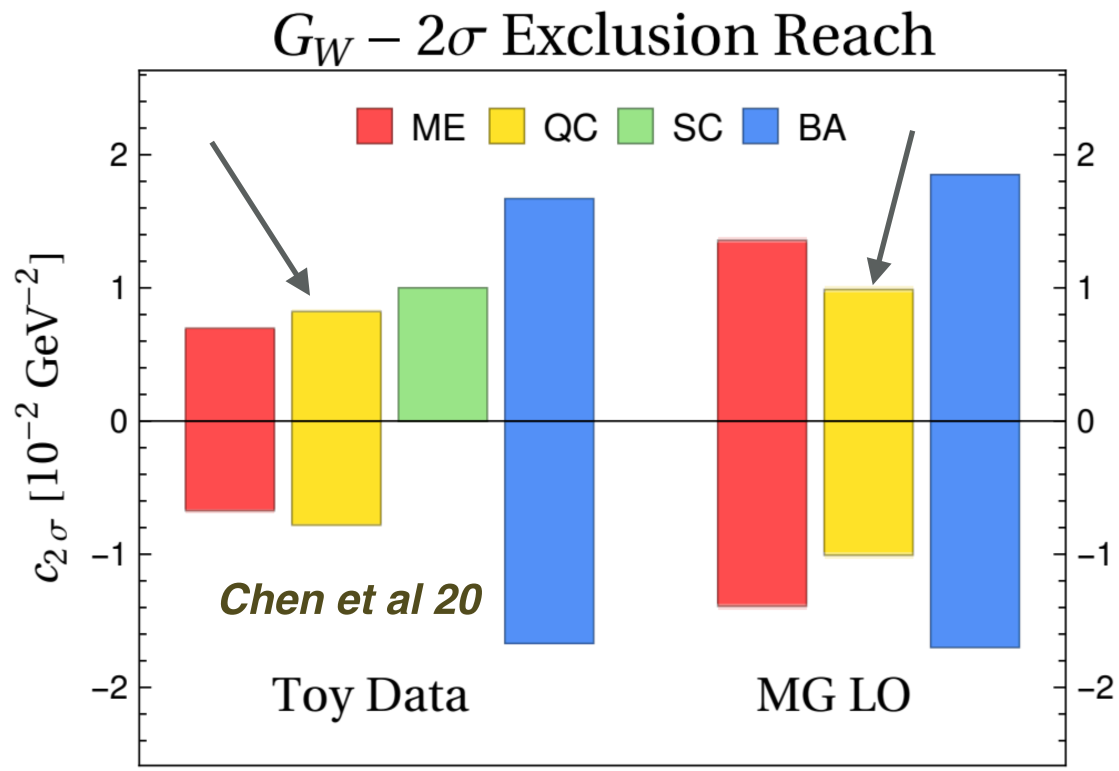
unbinned likelihoods, using our "convention"

$$\lambda(\mathcal{D}) \equiv \log \frac{\mathcal{L}(H_1|\mathcal{D})}{\mathcal{L}(H_0|\mathcal{D})} = \mathrm{N}(\mathrm{X}|H_0) - \mathrm{N}(\mathrm{X}|H_1) - \sum_{i=1}^{\mathcal{N}} \log \frac{d\sigma_0(x_i)}{d\sigma_1(x_i)} \,.$$

extended log-likehood ratio

Combined with **machine learning** methods, construct **optimally-sensitive observables** for EFT studies

How would these **unbinned likelihoods** be made public? This is a challenge for many reasons ….



also many great papers by Kyle and collaborators

Let's start making sure that the binned likelihoods are released