Wilks' theorem in EFT
CMS report

Brent R. Yates

brent.yates@cern.ch

Issue with EFT and Wilks' theorem

Quadratic terms in EFT fits can cause issues

Certain WC values cannot be explored because c^2 is positive, no longer lie within the parameter space

Single bin examples

Linear only

Wilks' theorem is valid

Quadratic dominant

Wilks' theorem can overcover

Quadratic and linear comparable

Wilks' theorem can undercover

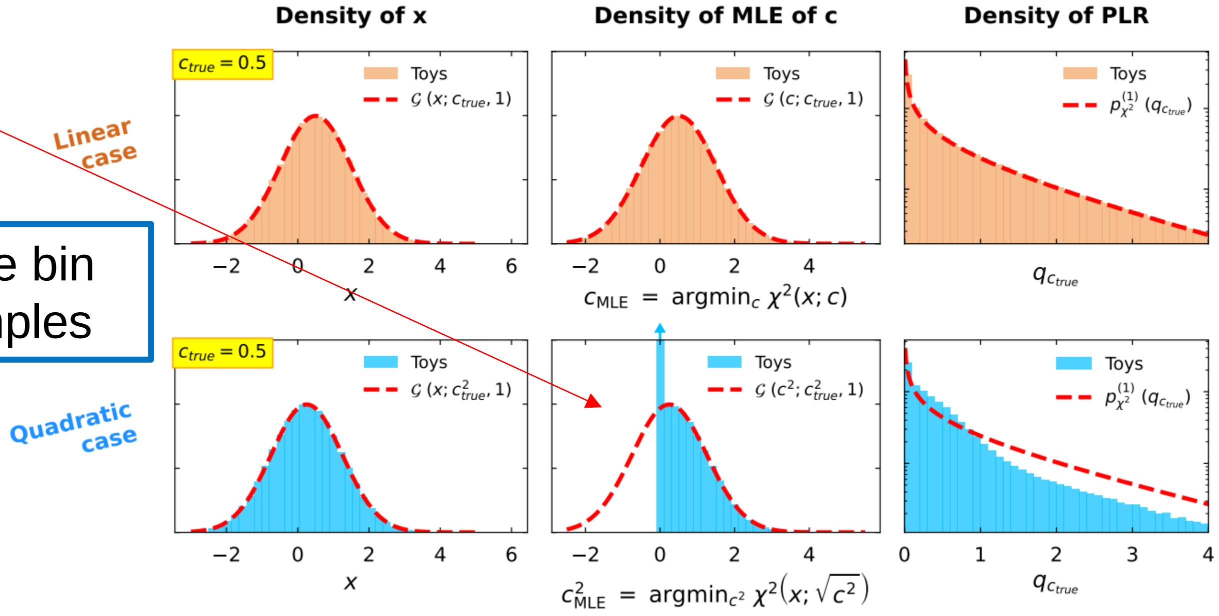


Figure 1: Visualisation of how Wilks' theorem is violated when a parameterisation cannot fit data fluctuations which fall beyond some boundary. This behaviour is observed in quadratic EFT fits (bottom row), which cannot access data below the SM hypothesis.

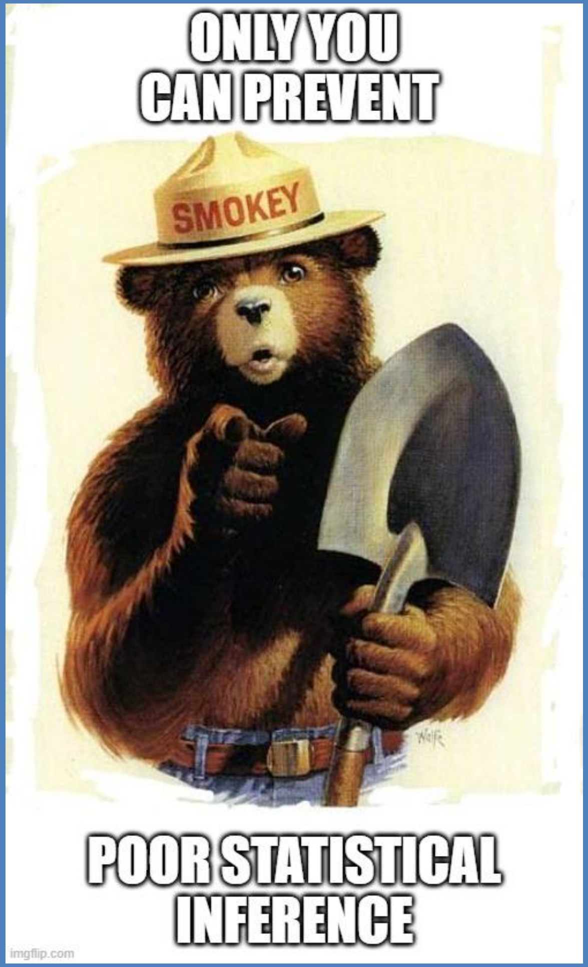
Wilks' Theorem

Theorem – *As the sample size approaches ∞, the distribution of the test statistic ($-2 * \ln$ likelihood ratio) asymptotically approaches the χ^2 distribution under the null hypothesis*

$$\lim_{N \to \infty} -2\ln(\Lambda) \sim \chi^2_{ndf}$$

$$\Lambda = \frac{\text{likelihood null model}}{\text{likelihood alterative modle}}$$

"Wilks' theorem assumes that the 'true' but unknown values of the estimated parameters lie within the interior of the supported parameter space"



Possible solutions

Possible solutions

The LHC EFT WG can simply agree to continue using Wilks' theorem

Pros

- Simplest path forward
- Consistent analysis results with previous publications

Cons

- Not fully correct; ignores issues presented today
- Potentially rule out new physics before $n\,\sigma$ is reached

Possible solutions

Each analysis can do a case-by-case exploration of their WCs

1. In an analysis fitting *only* linear components, Wilks' theorem is perfectly valid

2. If the WCs are quadratically dominated, Wilks' theorem will report an interval *larger* than it needs to be → a conservative estimate

3. If the WCs are in a region where the linear *and* quadratic terms are comparable, the interval *could* be smaller than it needs to be
   - If the test statistic q is sufficiently "stable" (i.e., not a large change in the shape) at the best fit point and the limits, it *may* be safe to use toys to extract the CIs

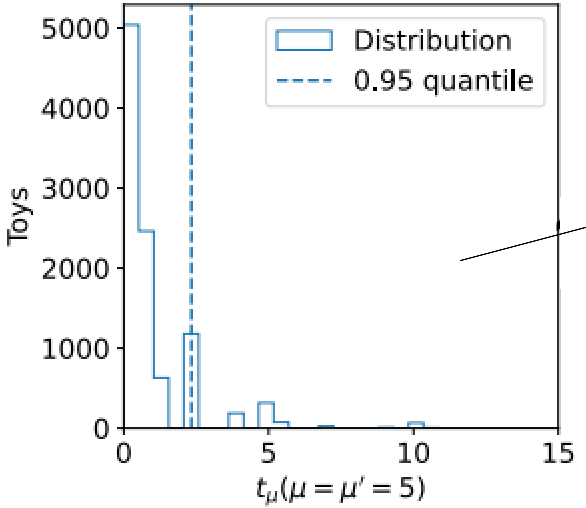
4. Estimate confidence intervals based on toys

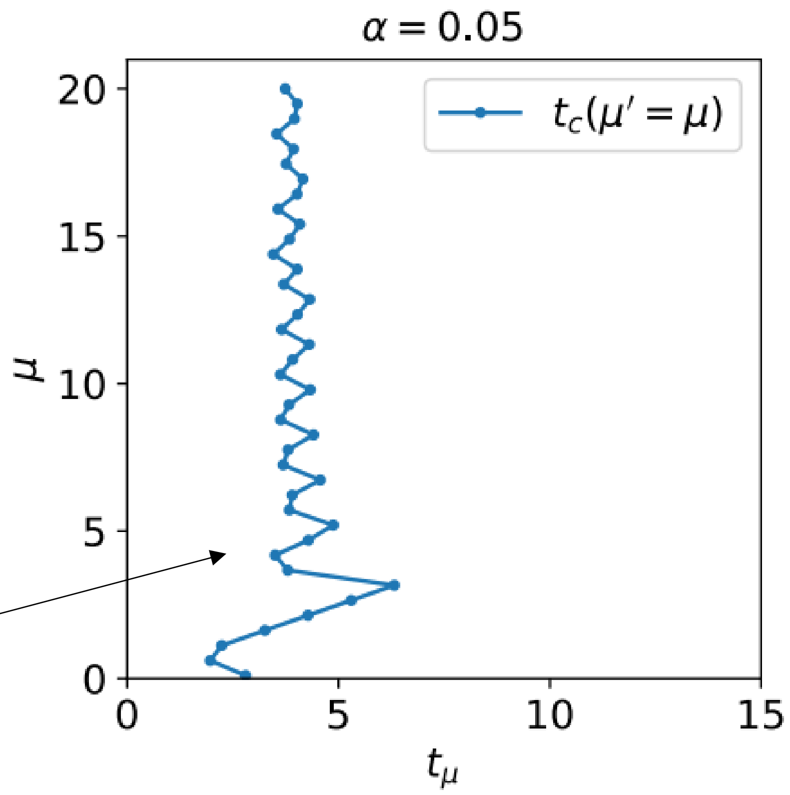
Please contact your experiment's statistics committee

Neyman interval

https://indico.cern.ch/event/1383972/contributions/5825309/attachments/2882605/5050900/CombineTutorial_Stats.pdf

Use toys to find the 68% and 95% confidence levels



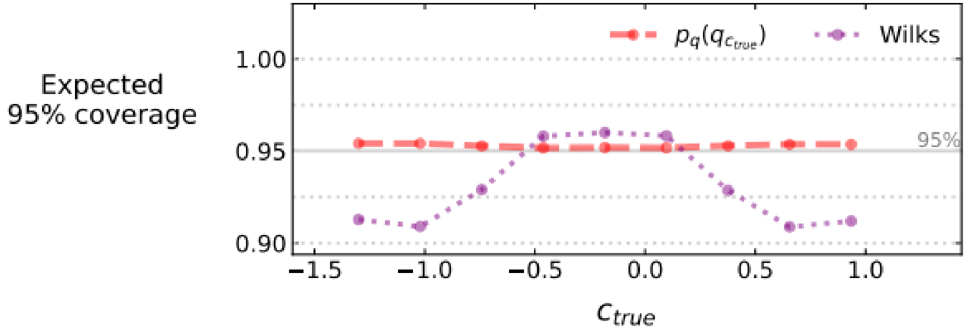


$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Updating software

The paper in [arXiv:2207.01350]() gives an examples and algorithm to compute the correct intervals

This could be incorporated into statistical packages (combine, phyhf, etc.)



5.2.8 The final algorithm

The final method for computing the PDF $p_q(q_{c_{\text{true}}})$ is summarised as follows: use Gaussian quadrature summation to compute the integral of Eq 88 over an arbitrarily wide interval of e.g. $\mu_{\Delta_0} - 5\sigma_{\Delta_0} \leq \Delta_0 \leq \mu_{\Delta_0} + 5\sigma_{\Delta_0}$, where the set "$\Delta_1 | q_{c_{\text{true}}}, \Delta_0$" is obtained using Algorithm 1. The Jacobian factor is calculated as **the reciprocal of** $\frac{dq_{c_{\text{true}}}}{d\Delta_1}$, which is obtained by applying the `JAX` auto-differentiation package to the forwards transformation.

Algorithm 1 Identify all values of Δ_1 consistent with a given $(q_{c_{\text{true}}}, \Delta_0)$-pair

Input: $(q_{c_{\text{true}}}, \Delta_0)$ and configurable constants $n_\sigma \sim 5$, $s_{\text{min}} \sim 5e-3$ and $\delta_{\text{min}} \sim 1e-3$
Returns: List of Δ_1 values of length 0, 1 or 2

```
Create empty list
for all Δ1^hyp in a broad linear scan between μ_Δ1 − n_σ σ_Δ1 and μ_Δ1 + n_σ σ_Δ1 do
    for first two c_MLE solutions of quartic equation q_ctrue(c_MLE, Δ0, Δ1^hyp) do
        Obtain Δ1^reco using inverse transformation of (c_MLE, Δ0)
        Calculate score s ← |Δ1^hyp − Δ1^reco|
        Require: s is in a local minimum (refine Δ1^hyp using iMinuit optimisation)
        Require: s < s_min
        Require: Δ1^hyp solution satisfies Class A, B, C(c0) or C(c1) conditions
        Require: |Δ1^hyp − Δ1^i| > δ_min · σ_Δ1 for all Δ1^i already in list
        Add Δ1^hyp to list
    end for
end for
Return list
```

Combined effective field theory interpretation of Higgs boson, electroweak vector boson, top quark, and multi-jet measurements

CMS-SMP-24-003

List of operators

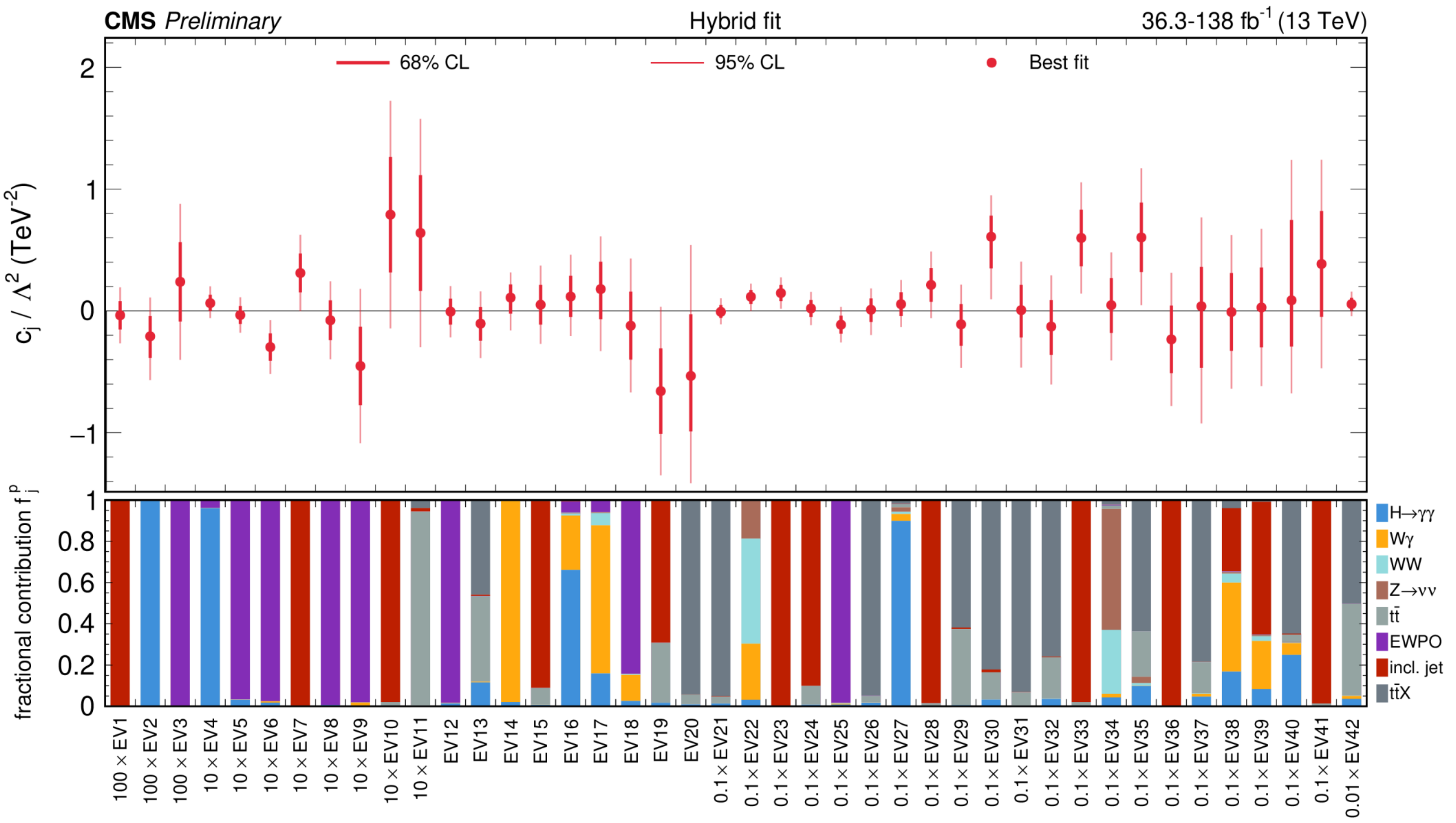
X^3		
$\mathcal{Q}_{G} = f^{abc} G_\mu^{a\nu} G_\nu^{b\rho} G_\rho^{c\mu}$	$\mathcal{Q}_{W} = \varepsilon^{ijk} W_\mu^{i\nu} W_\nu^{j\rho} W_\rho^{k\mu}$	
$H^4 D^2$		
$\mathcal{Q}_{H\Box} = (H^\dagger H)\Box(H^\dagger H)$	$\mathcal{Q}_{HD} = (D^\mu H^\dagger H)(H^\dagger D_\mu H)$	
$X^2 H^2$		
$\mathcal{Q}_{HG} = H^\dagger H G^a_{\mu\nu} G^{a\mu\nu}$	$\mathcal{Q}_{HW} = H^\dagger H W^i_{\mu\nu} W^{i\mu\nu}$	$\mathcal{Q}_{HB} = H^\dagger H B_{\mu\nu} B^{\mu\nu}$
$\mathcal{Q}_{HWB} = H^\dagger H W^i_{\mu\nu} B^{\mu\nu}$		
$\psi^2 H^3$		
$\mathcal{Q}_{tH} = (H^\dagger H)(\overline{Q}\tilde{H}t)$	$\mathcal{Q}_{bH} = (H^\dagger H)(\overline{Q}Hb)$	
$\psi^2 X H$		
$\mathcal{Q}_{tW} = (\overline{Q}\sigma^{\mu\nu}t)\sigma^i\tilde{H}W^i_{\mu\nu}$	$\mathcal{Q}_{tB} = (\overline{Q}\sigma^{\mu\nu}t)\tilde{H}B_{\mu\nu}$	$\mathcal{Q}_{tG} = (\overline{Q}\sigma^{\mu\nu}T^a t)\tilde{H}G^a_{\mu\nu}$
$\psi^2 H^2 D$		
$\mathcal{Q}^{(1)}_{Hl} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{l}_p\gamma^\mu l_r)$	$\mathcal{Q}^{(3)}_{Hl} = (H^\dagger i\overleftrightarrow{D}^i_\mu H)(\bar{l}_p\sigma^i\gamma^\mu l_r)$	$\mathcal{Q}_{He} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{e}_p\gamma^\mu e_r)$
$\mathcal{Q}^{(1)}_{Hq} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{q}\gamma^\mu q)$	$\mathcal{Q}^{(3)}_{Hq} = (H^\dagger i\overleftrightarrow{D}^i_\mu H)(\bar{q}\sigma^i\gamma^\mu q)$	$\mathcal{Q}_{Hu} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{u}\gamma^\mu u)$
$\mathcal{Q}_{Hd} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{d}\gamma^\mu d)$	$\mathcal{Q}^{(1)}_{HQ} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{Q}\gamma^\mu Q)$	$\mathcal{Q}^{(3)}_{HQ} = (H^\dagger i\overleftrightarrow{D}^i_\mu H)(\overline{Q}\sigma^i\gamma^\mu Q)$
$\mathcal{Q}_{Ht} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{t}\gamma^\mu t)$	$\mathcal{Q}_{Hb} = (H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{b}\gamma^\mu b)$	

$\psi^4, (\bar{L}L)(\bar{L}L)$		
$\mathcal{Q}^{(1)}_{lq} = (\bar{l}_p\gamma_\mu l_r)(\bar{q}\gamma^\mu q)$	$\mathcal{Q}^{(3)}_{lq} = (\bar{l}_p\sigma^i\gamma_\mu l_r)(\bar{q}\sigma^i\gamma^\mu q)$	$\mathcal{Q}^{(1)}_{lQ} = (\bar{l}_p\gamma_\mu l_r)(\overline{Q}\gamma^\mu Q)$
$\mathcal{Q}^{(3)}_{lQ} = (\bar{l}_p\sigma^i\gamma_\mu l_r)(\overline{Q}\sigma^i\gamma^\mu Q)$	$\mathcal{Q}^{(1)}_{QQ} = (\overline{Q}\gamma_\mu Q)(\overline{Q}\gamma^\mu Q)$	$\mathcal{Q}_{ll} = (\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$
$\mathcal{Q}^{(1,1)}_{qq} = (\bar{q}\gamma_\mu q)(\bar{q}\gamma^\mu q)$	$\mathcal{Q}^{(1,8)}_{qq} = (\bar{q}T^a\gamma_\mu q)(\bar{q}T^a\gamma^\mu q)$	$\mathcal{Q}^{(3,1)}_{qq} = (\bar{q}\sigma^i\gamma_\mu q)(\bar{q}\sigma^i\gamma^\mu q)$
$\mathcal{Q}^{(3,8)}_{qq} = (\bar{q}\sigma^i T^a\gamma_\mu q)(\bar{q}\sigma^i T^a\gamma^\mu q)$	$\mathcal{Q}^{(1,1)}_{Qq} = (\overline{Q}\gamma_\mu Q)(\bar{q}\gamma^\mu q)$	$\mathcal{Q}^{(1,8)}_{Qq} = (\overline{Q}T^a\gamma_\mu Q)(\bar{q}T^a\gamma^\mu q)$
$\mathcal{Q}^{(3,1)}_{Qq} = (\overline{Q}\sigma^i\gamma_\mu Q)(\bar{q}\sigma^i\gamma^\mu q)$	$\mathcal{Q}^{(3,8)}_{Qq} = (\overline{Q}\sigma^i T^a\gamma_\mu Q)(\bar{q}\sigma^i T^a\gamma^\mu q)$	
$\psi^4, (\bar{R}R)(\bar{R}R)$		
$\mathcal{Q}_{et} = (\bar{e}_p\gamma_\mu e_r)(\bar{t}\gamma^\mu t)$	$\mathcal{Q}_{tt} = (\bar{t}\gamma_\mu t)(\bar{t}\gamma^\mu t)$	$\mathcal{Q}^{(1)}_{uu} = (\bar{u}\gamma_\mu u)(\bar{u}\gamma^\mu u)$
$\mathcal{Q}^{(8)}_{uu} = (\bar{u}T^a\gamma_\mu u)(\bar{u}T^a\gamma^\mu u)$	$\mathcal{Q}^{(1)}_{tu} = (\bar{t}\gamma_\mu t)(\bar{u}\gamma^\mu u)$	$\mathcal{Q}^{(8)}_{tu} = (\bar{t}T^a\gamma_\mu t)(\bar{u}T^a\gamma^\mu u)$
$\mathcal{Q}^{(1)}_{dd} = (\bar{d}\gamma_\mu d)(\bar{d}\gamma^\mu d)$	$\mathcal{Q}^{(8)}_{dd} = (\bar{d}T^a\gamma_\mu d)(\bar{d}T^a\gamma^\mu d)$	$\mathcal{Q}^{(1)}_{ud} = (\bar{u}\gamma_\mu u)(\bar{d}\gamma^\mu d)$
$\mathcal{Q}^{(8)}_{ud} = (\bar{u}T^a\gamma_\mu u)(\bar{d}T^a\gamma^\mu d)$	$\mathcal{Q}^{(1)}_{td} = (\bar{t}\gamma_\mu t)(\bar{d}\gamma^\mu d)$	$\mathcal{Q}^{(8)}_{td} = (\bar{t}T^a\gamma_\mu t)(\bar{d}T^a\gamma^\mu d)$
$\psi^4, (\bar{L}L)(\bar{R}R)$		
$\mathcal{Q}_{lu} = (\bar{l}_p\gamma_\mu l_r)(\bar{u}\gamma^\mu u)$	$\mathcal{Q}_{lt} = (\bar{l}_p\gamma_\mu l_r)(\bar{t}\gamma^\mu t)$	$\mathcal{Q}^{(1)}_{qu} = (\bar{q}\gamma_\mu q)(\bar{u}\gamma^\mu u)$
$\mathcal{Q}^{(8)}_{qu} = (\bar{q}T^a\gamma_\mu q)(\bar{u}T^a\gamma^\mu u)$	$\mathcal{Q}^{(1)}_{Qu} = (\overline{Q}\gamma_\mu Q)(\bar{u}\gamma^\mu u)$	$\mathcal{Q}^{(8)}_{Qu} = (\overline{Q}T^a\gamma_\mu Q)(\bar{u}T^a\gamma^\mu u)$
$\mathcal{Q}^{(1)}_{qt} = (\bar{q}\gamma_\mu q)(\bar{t}\gamma^\mu t)$	$\mathcal{Q}^{(8)}_{qt} = (\bar{q}T^a\gamma_\mu q)(\bar{t}T^a\gamma^\mu t)$	$\mathcal{Q}^{(1)}_{Qt} = (\overline{Q}\gamma_\mu Q)(\bar{t}\gamma^\mu t)$
$\mathcal{Q}^{(8)}_{Qt} = (\overline{Q}T^a\gamma_\mu Q)(\bar{t}T^a\gamma^\mu t)$	$\mathcal{Q}^{(1)}_{qd} = (\bar{q}\gamma_\mu q)(\bar{d}\gamma^\mu d)$	$\mathcal{Q}^{(8)}_{qd} = (\bar{q}T^a\gamma_\mu q)(\bar{d}T^a\gamma^\mu d)$
$\mathcal{Q}^{(1)}_{Qd} = (\overline{Q}\gamma_\mu Q)(\bar{d}\gamma^\mu d)$	$\mathcal{Q}^{(8)}_{Qd} = (\overline{Q}T^a\gamma_\mu Q)(\bar{d}T^a\gamma^\mu d)$	

Analyses in combination

Analysis	Type of measurement	Observables used	Experimental likelihood
$\mathrm{H} \to \gamma\gamma$	Diff. cross sections	STXS bins [41]	✓
$\mathrm{W}\gamma$	Fid. diff. cross sections	$p_{\mathrm{T}}^{\gamma} \times \lvert\phi_f\rvert$	✓
WW	Fid. diff. cross sections	$m_{\ell\ell}$	✓
$\mathrm{Z} \to \nu\nu$	Fid. diff. cross sections	$p_{\mathrm{T}}^{\mathrm{Z}}$	✓
$\mathrm{t\bar{t}}$	Fid. diff. cross sections	$M_{\mathrm{t\bar{t}}}$	×
EWPO	Pseudo-observables	$\Gamma_{\mathrm{Z}},\ \sigma^0_{\mathrm{had}},\ R_\ell,\ R_c,\ R_b,\ A^{0,\ell}_{FB},\ A^{0,\mathrm{c}}_{FB},\ A^{0,\mathrm{b}}_{FB}$	×
Inclusive jet	Fid. diff. cross sections	$p_{\mathrm{T}}^{\mathrm{jet}} \times \lvert y^{\mathrm{jet}}\rvert$	×
$\mathrm{t\bar{t}X}$	Direct EFT	Yields in regions of interest	✓

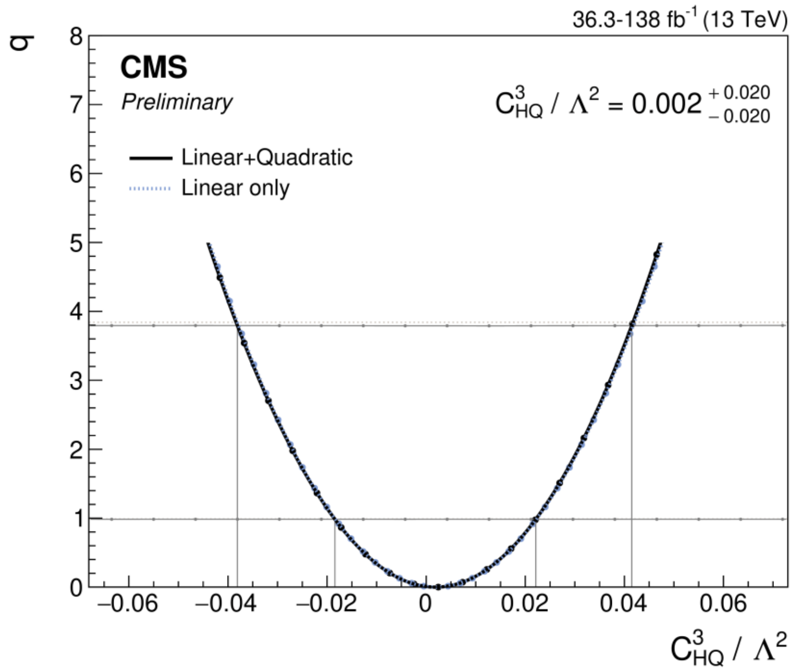
PCA fit



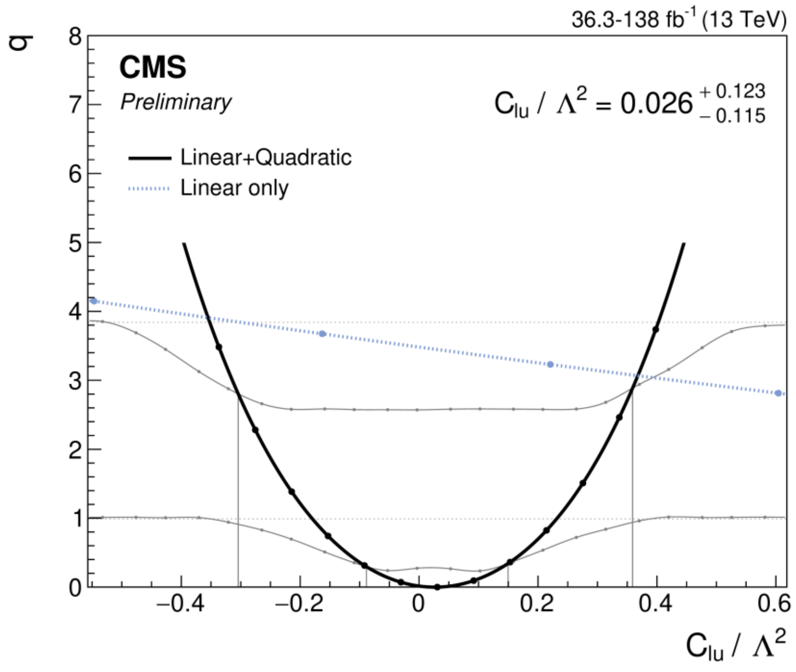
CMS Preliminary
Hybrid fit
36.3-138 fb^{-1} (13 TeV)
68% CL
95% CL
Best fit
c_j / Λ^2 (TeV^{-2})
fractional contribution f_j^p
H→γγ
Wγ
WW
Z→νν
$t\bar{t}$
EWPO
incl. jet
$t\bar{t}X$
100 × EV1
100 × EV2
100 × EV3
10 × EV4
10 × EV5
10 × EV6
10 × EV7
10 × EV8
10 × EV9
10 × EV10
10 × EV11
EV12
EV13
EV14
EV15
EV16
EV17
EV18
EV19
EV20
0.1 × EV21
0.1 × EV22
0.1 × EV23
0.1 × EV24
0.1 × EV25
0.1 × EV26
0.1 × EV27
0.1 × EV28
0.1 × EV29
0.1 × EV30
0.1 × EV31
0.1 × EV32
0.1 × EV33
0.1 × EV34
0.1 × EV35
0.1 × EV36
0.1 × EV37
0.1 × EV38
0.1 × EV39
0.1 × EV40
0.1 × EV41
0.01 × EV42

Intervals based on test statistic and toys

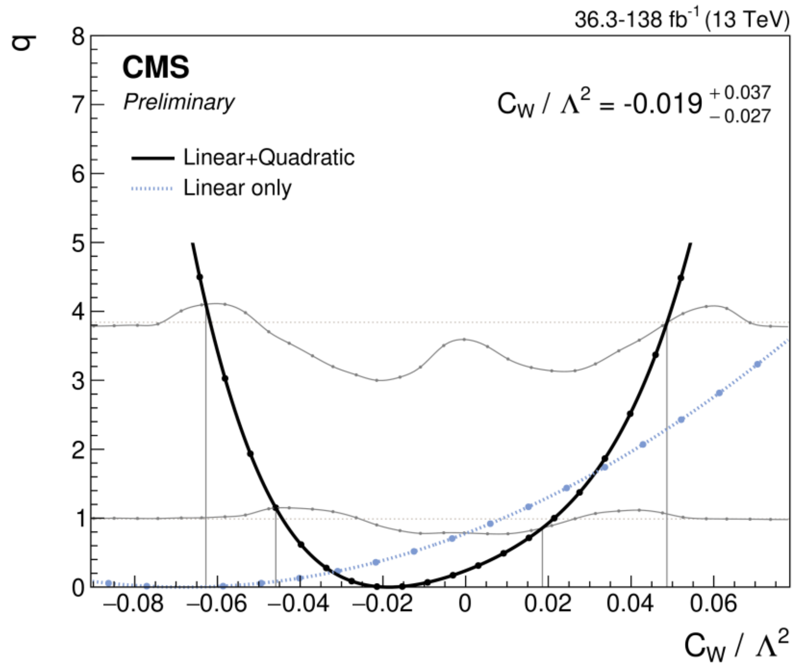
Linear dominates



Quadratic dominates

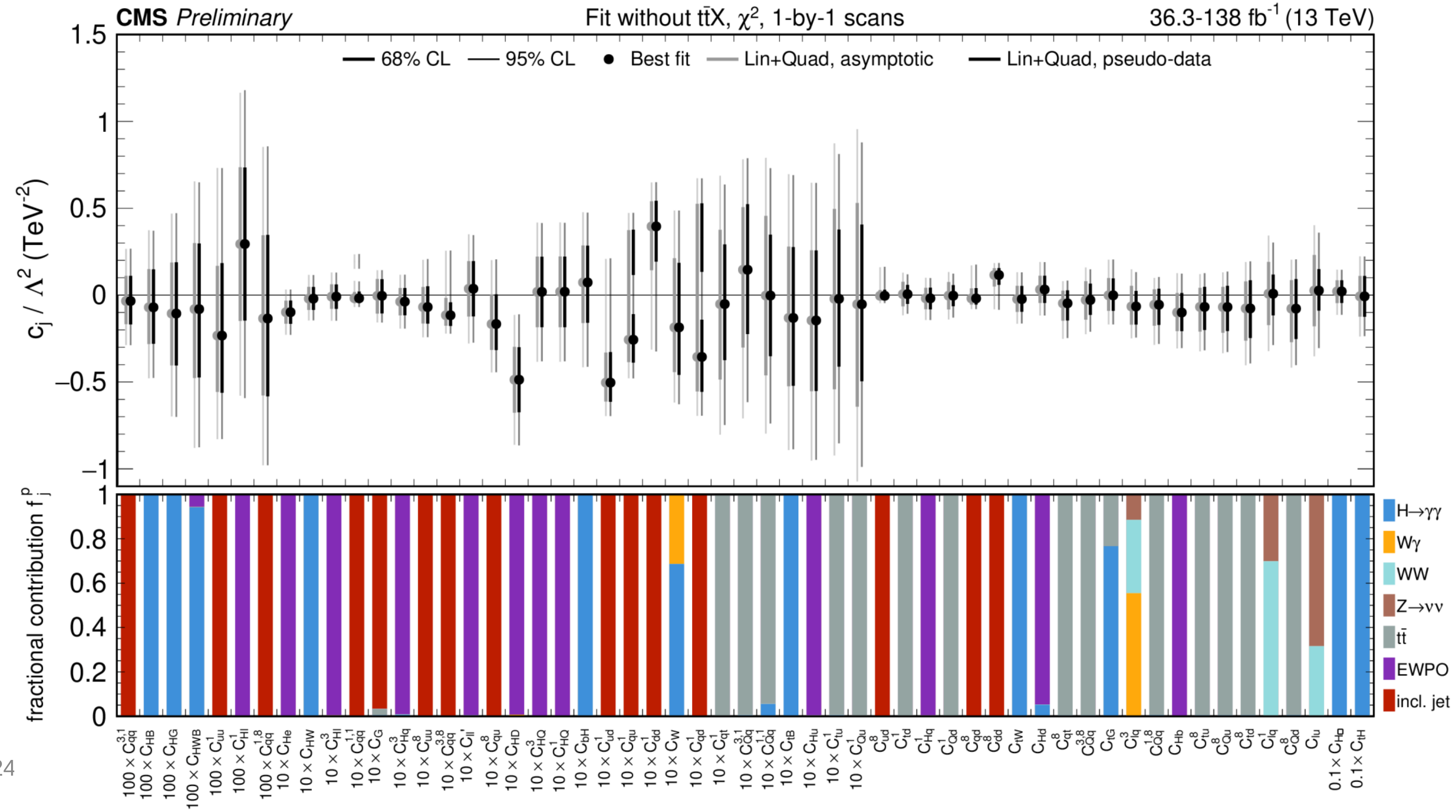


Linear + Quadratic



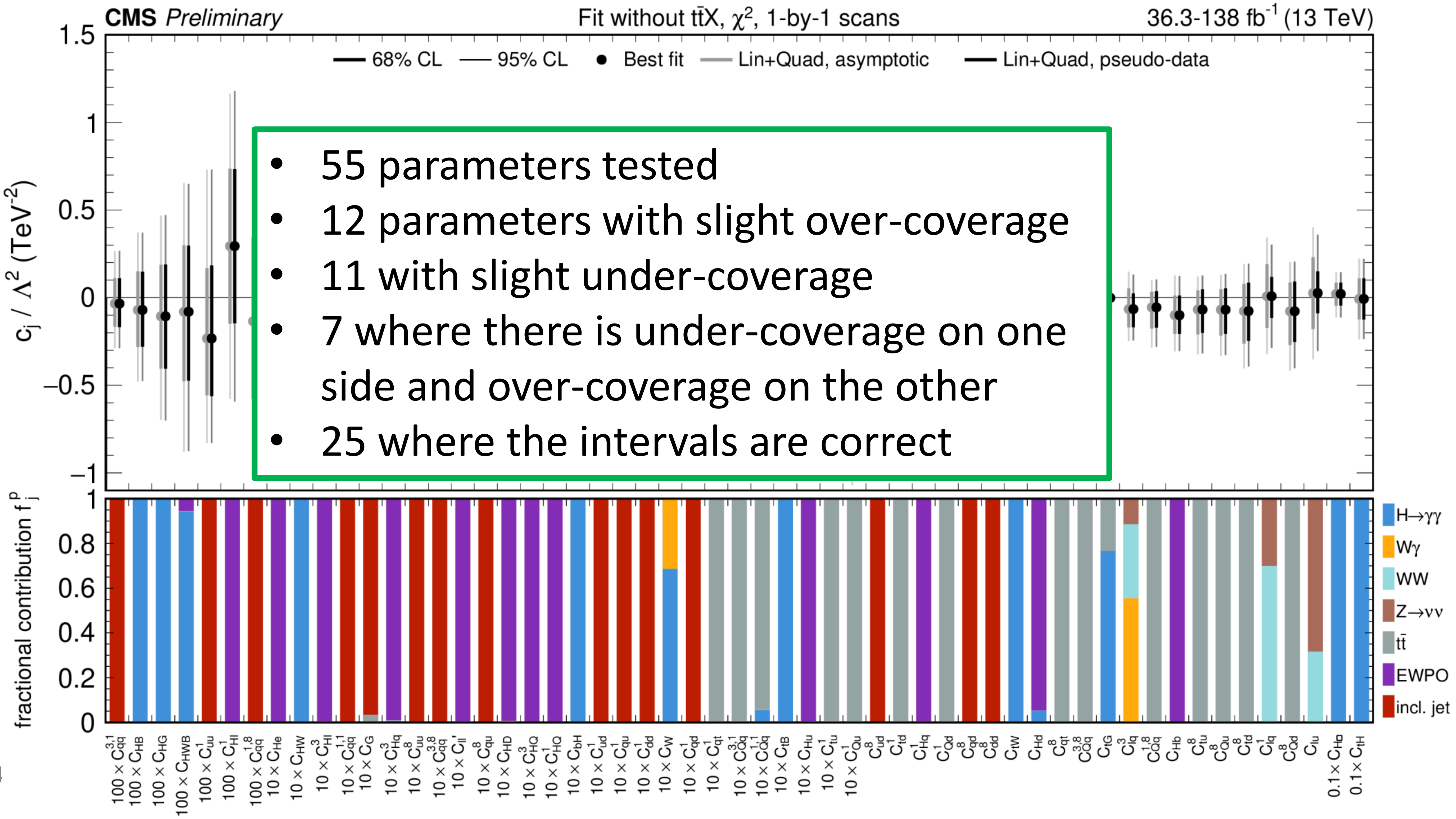
Lin+Quad, asymptotic   Lin+Quad, pseudo-data

Comparison of coverage in full fit



Lin+Quad, asymptotic   Lin+Quad, pseudo-data

Comparison of coverage in full fit



- 55 parameters tested
- 12 parameters with slight over-coverage
- 11 with slight under-coverage
- 7 where there is under-coverage on one side and over-coverage on the other
- 25 where the intervals are correct

Summary

Wilks' theorem is often used to extract asymptotic limits from likelihood ratios

Quadratic parameterizations of WCs can break the Wilks' theorem ($N \sim c^2 \rightarrow c > 0$)

Interval coverage is no longer guaranteed

CMS-SMP-24-003 has explored this issue using toys

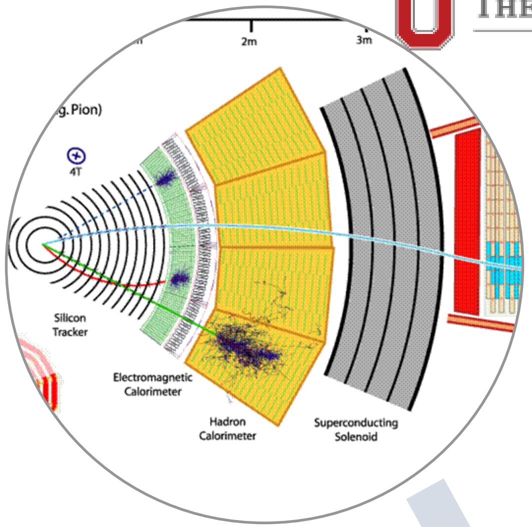
- Variations in coverage observed, with roughly half having correct coverage

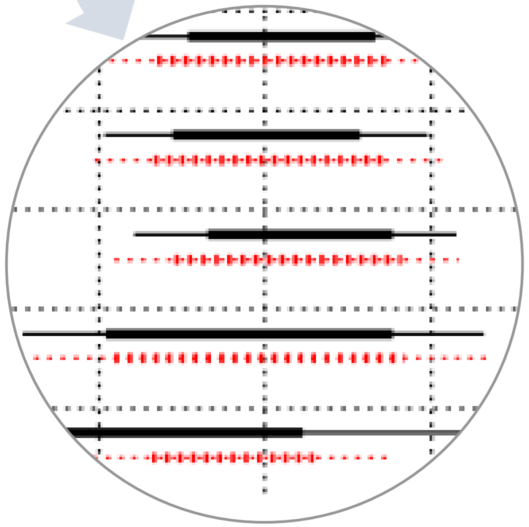
Backup

Overview

Most analyses measure some sort of rate or parameter

- Cross section σ
- Differential cross-section $\frac{d\sigma}{dp_{\mathrm{T}}}, \frac{d\sigma}{d|\eta|}, \frac{d\sigma}{d\cos\chi}$, etc.
- Top mass m_{t}
- Fragmentation shape r_{b}
- EFT Wilson coefficients c_i





We typically extract these quantities using statistical inference

- Provide templates and/or parametric functions that model the data
- Find best-fit values

The profiled likelihood ratio has become the most common method of inference at the LHC

Probability and likelihood

Probability of observing the data (x) given parameter(s) (θ)

$$p(x|\theta)$$

Likelihood is product of probabilities of parameter(s) *given* the data

$$\mathcal{L}(\theta) = p(\theta|x) = \prod_{\theta \epsilon \Theta} \frac{p(x|\theta)}{p(x)} * p(\theta)$$

Bayes theorem

where Θ is the entire parameter space

$p(x)$ is a normalization term and is typically ignored in the frequentist interpretation (when we compute $\max(\Delta \ln \mathcal{L}(\theta))$ the normalization isn't important)

Priors

Prior probabilities come from previous measurements of a certain quantity

- Usually nuisance parameters (NPs) in LHC analyses
- WCs have “flat priors” → no assumption is made on the possible values ($p(c) = 1$)

It is a penalty score in the likelihood to prevent the optimization step from pulling a particular parameter too far from its previously measured value

Example

We typically use a log-normal prior for the luminosity

A given CMS analysis is unlikely to measure the luminosity (e.g., Run 2 $138\ \mathrm{fb}^{-1}$) more accurately than the official delivered/recorded values

We allow the luminosity to vary *slightly* to accommodate correlations in the analysis, and the log-normal term prevents it from going below 0

The profiled likelihood ratio

Easier to minimize $-\ln \mathcal{L}$

In general, the test statistic is defined as

$$q = -2\ln\frac{\mathcal{L}(c_i, \hat{\hat{\theta}})}{\mathcal{L}(\hat{c}_i, \hat{\theta})}$$

where $\hat{c}_i$ and $\hat{\theta}$ minimizes $-\ln \mathcal{L}$ ("best fit point" or MLE), and $\hat{\hat{\theta}}$ minimizes $-\ln \mathcal{L}$ for a particular value of c_i (often the SM $c_i = 0$)

Statistical inference using the test statistic

The test statistic q is a measure of how compatible the data is with the model at some reference point (e.g., the SM)

The value(s) of the WCs c_i which minimize q are know as the "best fit values"

To give an uncertainty on the values, we typically use Wilks' theorem to extract confidence intervals around the best fit value

- Often quoted as 68, 95, and 99.7% (or 1σ, 2σ, and 3σ for 1D fits)
- The intervals are $q = 1,4,9$ for 1D, $q = 2.30, 6.18, 11.83$ for 2D, etc.

$$q = -2\ln\frac{\mathcal{L}(c_i, \hat{\hat{\theta}})}{\mathcal{L}(\hat{c}_i, \hat{\theta})}$$

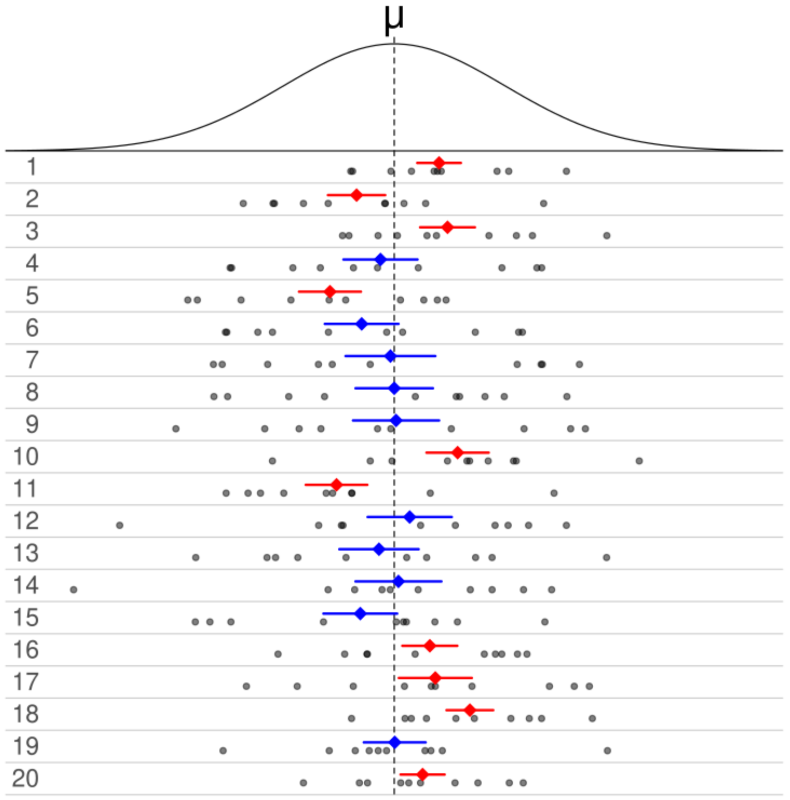
Confidence intervals

In the frequentist interpretation, the confidence interval is a statement about *repeated* experiments

<u>Example</u>

A confidence interval of 95% means if the *same* experiment were repeated *a large number of times* ($N \rightarrow \infty$), 95% of those intervals would contain the true value

It is ***not*** a statement on the probability of observing a particular result



https://en.wikipedia.org/wiki/Confidence_interval

Example from TOP-19-001

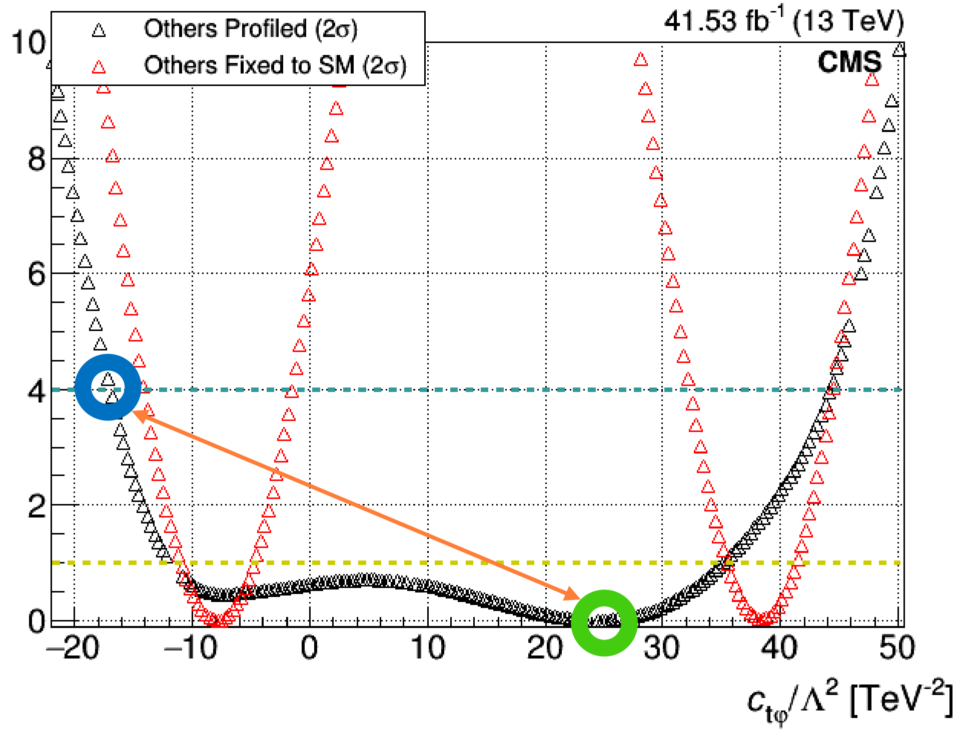
Wilks' theorem states $-2\Delta\ln\left(\mathcal{L}\left(\vec{c},\vec{\theta}\right)\right)\xrightarrow{N\to\infty}\chi^2_{ndf}$

Degrees of freedom – Difference between number of parameters in the best fit and the evaluated point

Example ($c_{\mathrm{t}\varphi}$)

In a **1D** profiled scans, we compare the best fit point (16 WCs profiled) to a specific value of $c_{\mathrm{t}\varphi}$ (15 WCs profiled), giving **1 degree of freedom**

Likelihood scan
Set $c_{\mathrm{t}\varphi}$ to specific values
Profile the other WCs or **fix** to 0
Profile all NPs



Rotations for PCA