Introduction to Combine (part 2)

Susan Dittmer, Northwestern University

CMS Topical Workshop on Off-shell Higgs
Boson Production at LPC

March 28, 2024

Likelihood Scans

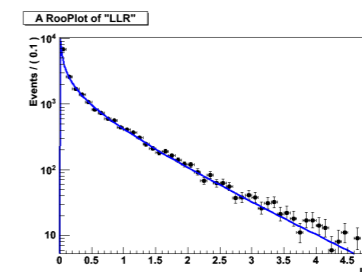- Recall Wilks' theorem:

$$p(t_\mu) \approx \chi^2(dim[\mu])$$
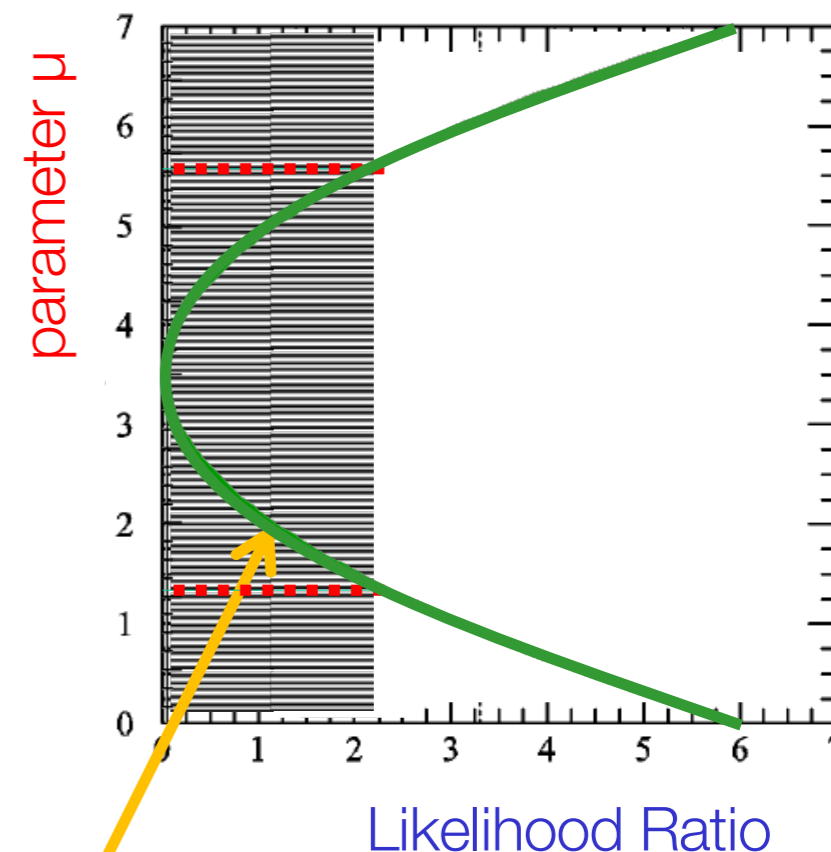
- where

$$t_\mu(x_{obs}, \mu) = -2ln(\lambda_p(\mu))$$

$$\lambda_p(\mu) = \frac{L(\mu, \hat{\hat{\nu}}(\mu))}{L(\hat{\mu}, \hat{\nu})}$$
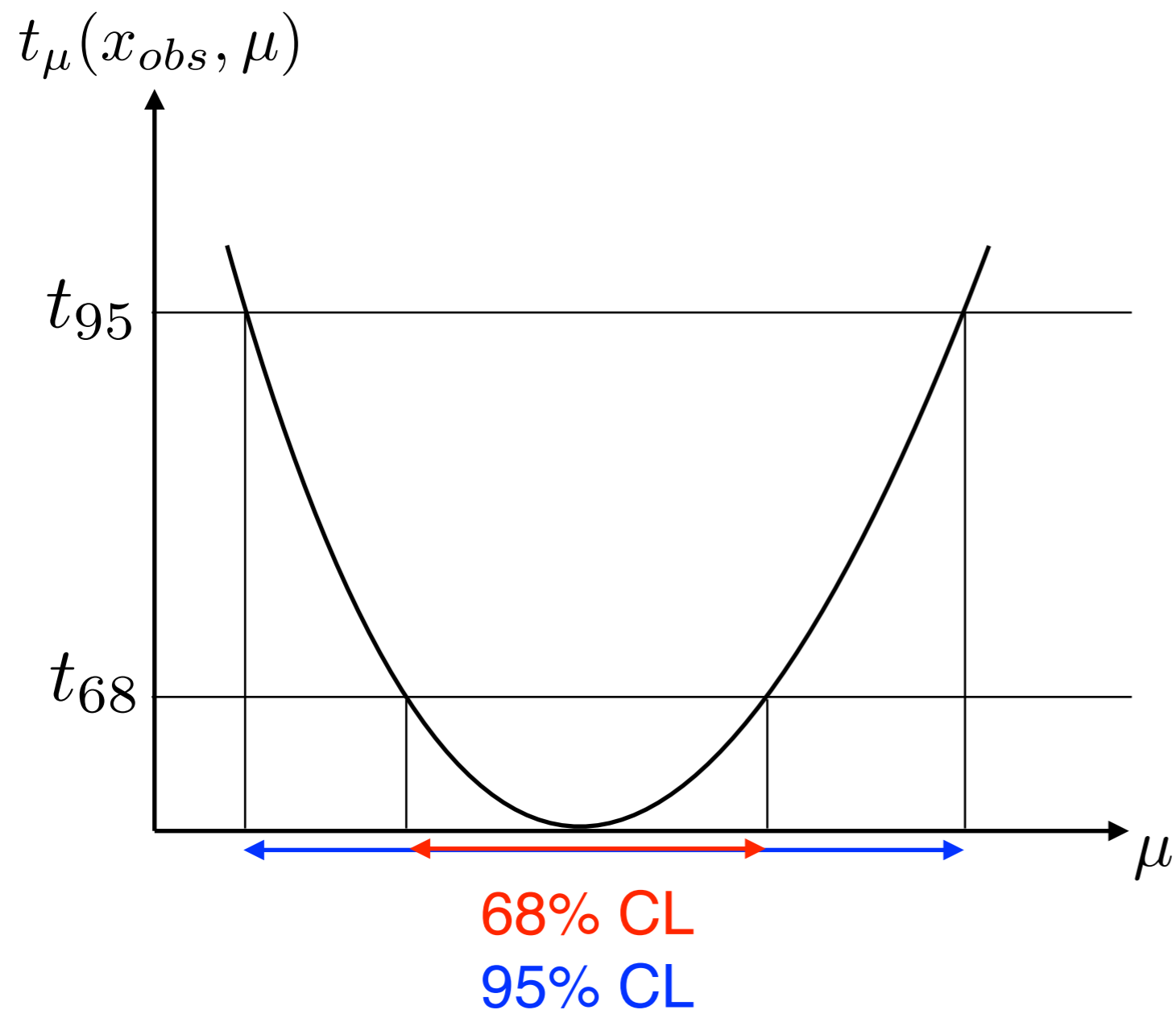
Max likelihood for a given μ
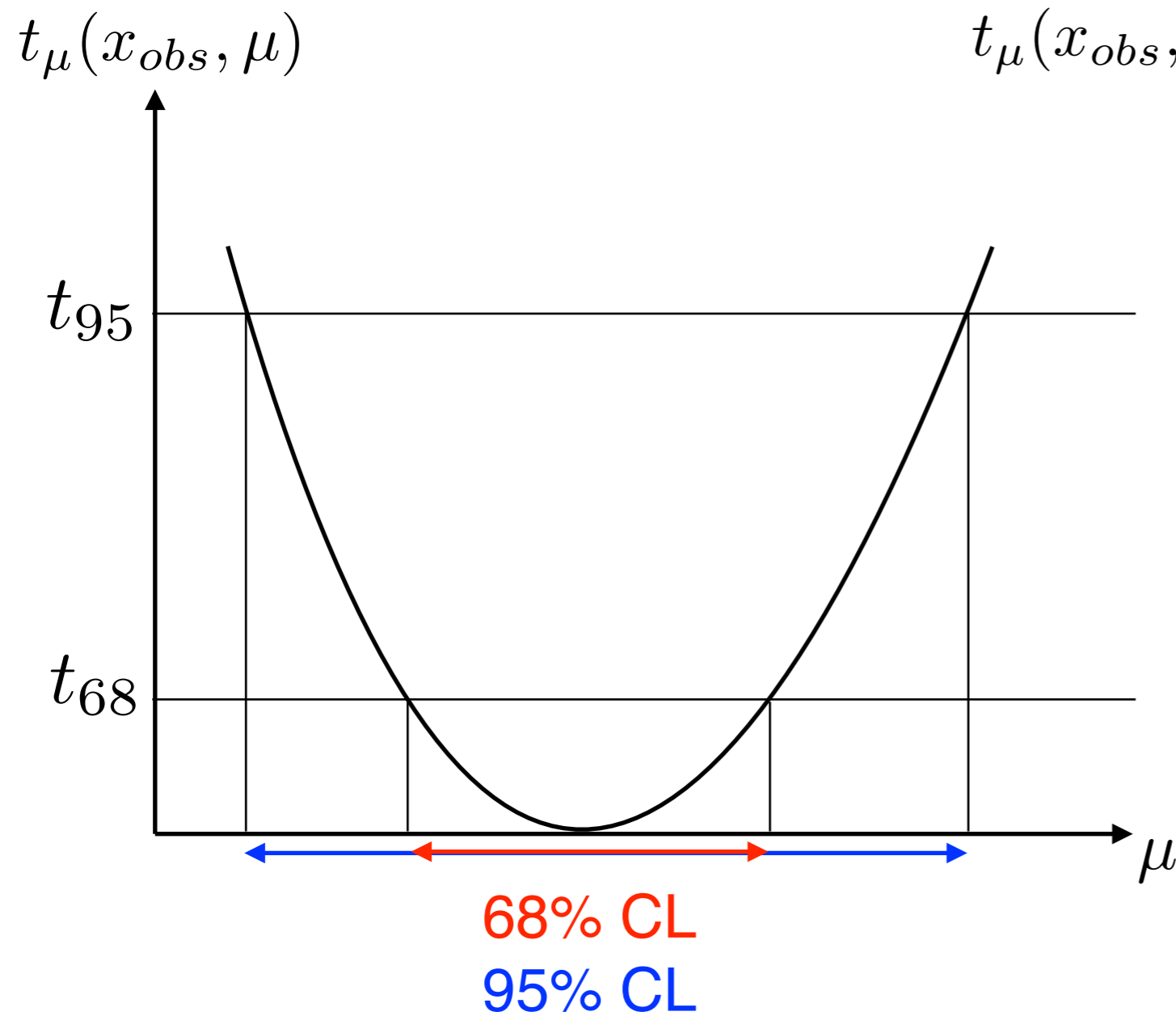
Maximum likelihood

$t_\mu(x,\mu)$

parameter μ

Likelihood Ratio

Measurement = $t_\mu(x_{obs}, \mu)$ is now a function of μ

Likelihood Scans
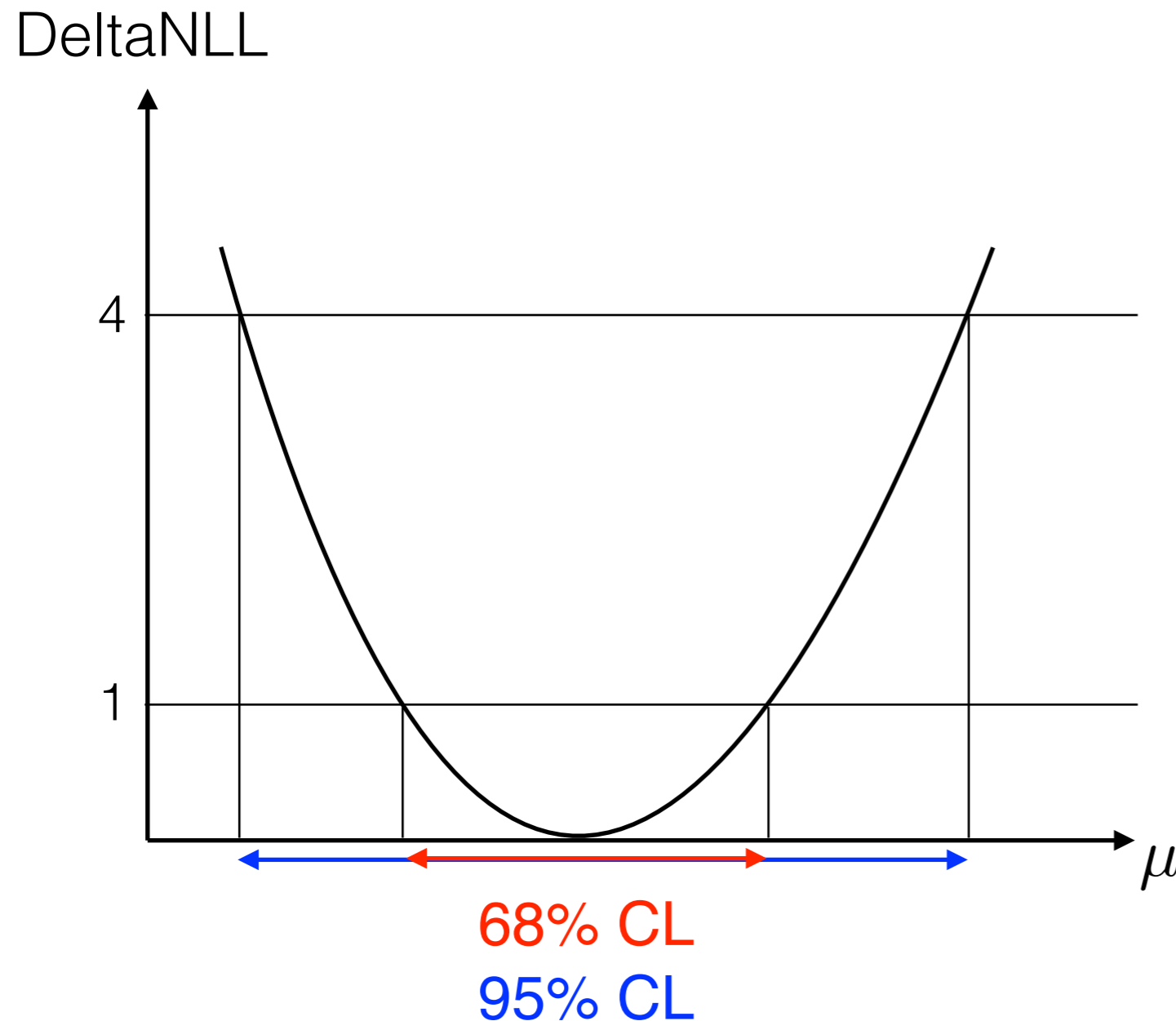
Likelihood Scans

$$t_\mu(x_{obs}, \mu) = -2ln(\lambda_p(\mu))$$

$$= -2ln(\frac{L(\mu, \hat{\hat{\nu}}(\mu))}{L(\hat{\mu}, \hat{nu})})$$

$$= NLL(\mu, \hat{\hat{\nu}}(\mu)) - NLL(\hat{\mu}, \hat{\nu})$$

68% CL

95% CL

Likelihood Scans

DeltaNLL

t_{68} and t_{95} give one-sided intervals with 68% and 95% coverage on a 1D χ^2

Recall χ^2 definition:

'distribution of a sum of the squares of k independent standard normal random variables'

68(95)% actually means '1(2) sigma'
—> $t_{68(95)} = 1(4)$

Setup

We will use CombineTools for plotting, so do a sparse checkout from CombineHarvester:

```
cmssw-el7
cd CMSSW_11_3_4/src
bash <(curl -s https://raw.githubusercontent.com/cms-analysis/
CombineHarvester/main/CombineTools/scripts/sparse-checkout-ssh.sh)
scramv1 b
```

CombineTools has a number of useful scripts: plotting, datacard validation, grid submission, etc…

Reuse the H—>ττ data card from before, but now produce workspace with inclusive signal strength:

```
cd data/tutorials/htt/125
text2workspace.py --mass 125 htt_tt.txt -o ../../../../htt_tt_incl.root
cd ../../../../
```

Likelihood Scan

Then do likelihood scan:

```
combine -M MultiDimFit --mass 125 -n incl \
        --algo grid --points 100 --autoRange 2 \
        -d htt_tt_incl.root
```

—autoRange 2 says to scan over range ±2σ in r

What is happening?

- Do initial fit to get best fit values of parameters, plus minimum NLL

- Scan 100 even spaced steps in r over range (r0-2σ,r0+2σ)

  - Compute profile likelihood — max likelihood for given r

- Store r, deltaNLL (+ a few other variables) for each point in output tree

Plot Likelihood Scan

```
python ../../CombineHarvester/CombineTools/scripts/plot1DScan.py \
higgsCombineincl.MultiDimFit.mH125.root
```
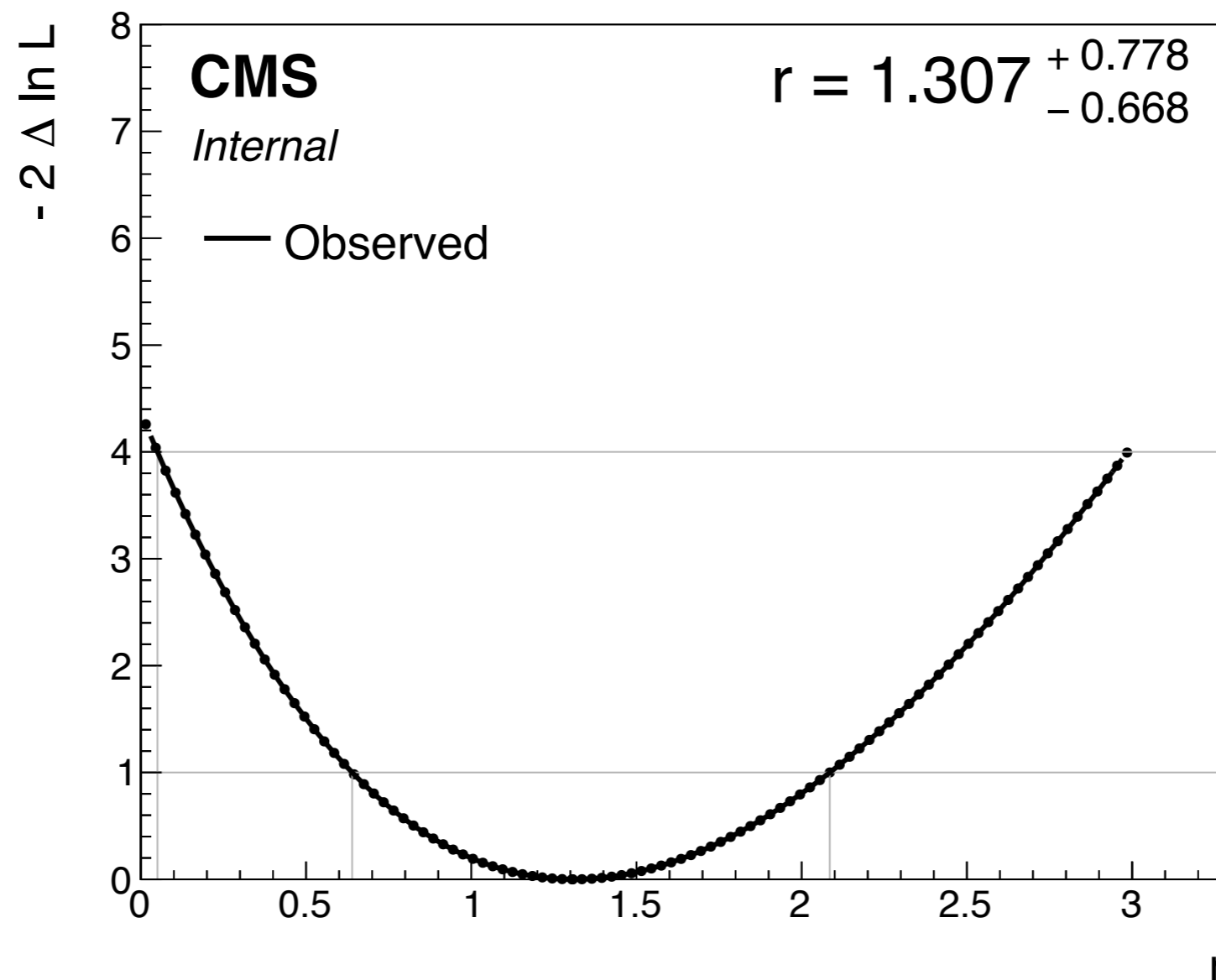
Stitching Likelihood Scans

May want to get more scan points in a given region — how to do this?
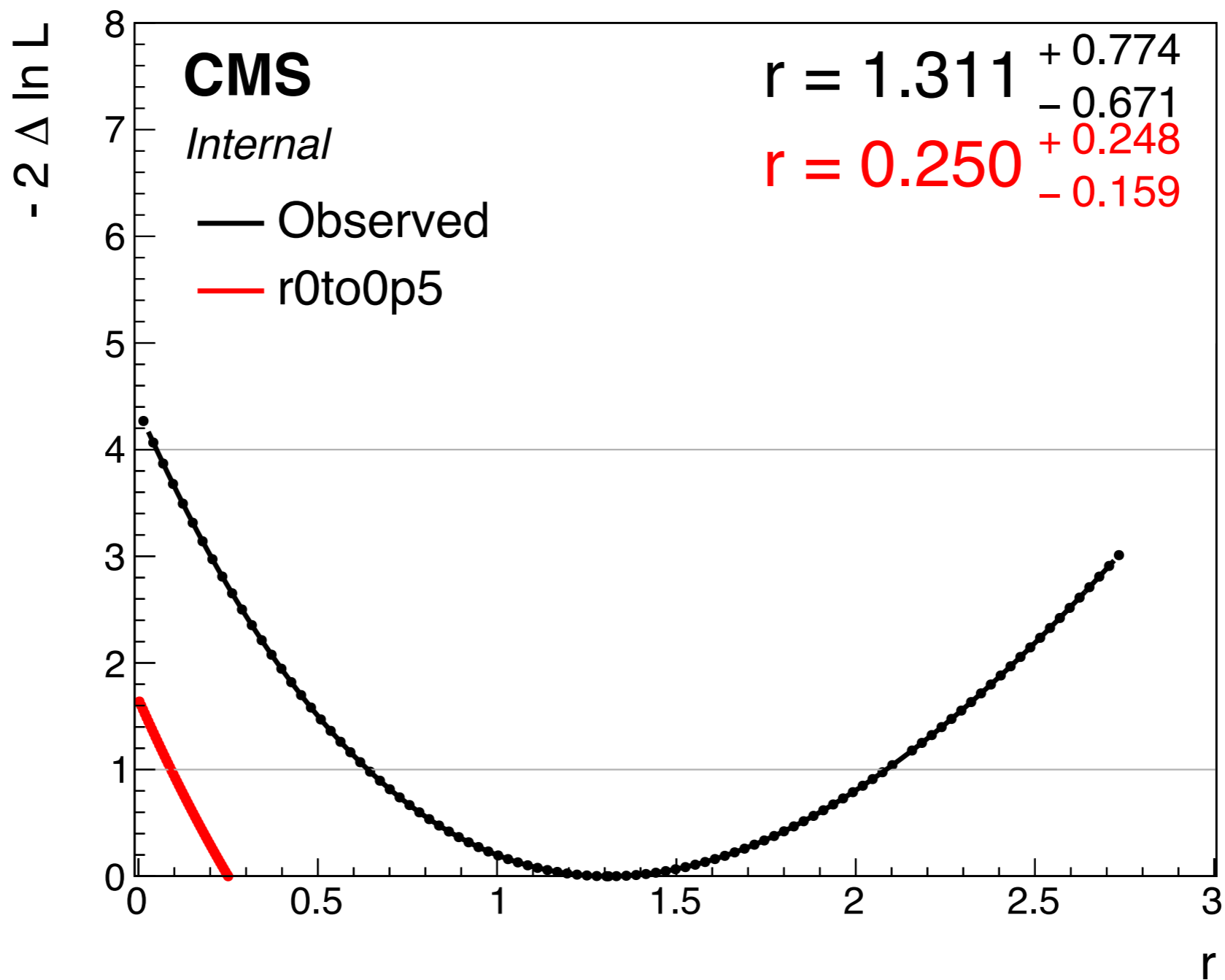
Try re-running scan with smaller range:

```
combine -M MultiDimFit --mass 125 -n r0to0p5 \
        --algo grid --points 100 --setParameterRange r=0,0.5 \
        -d htt_tt_incl.root
```
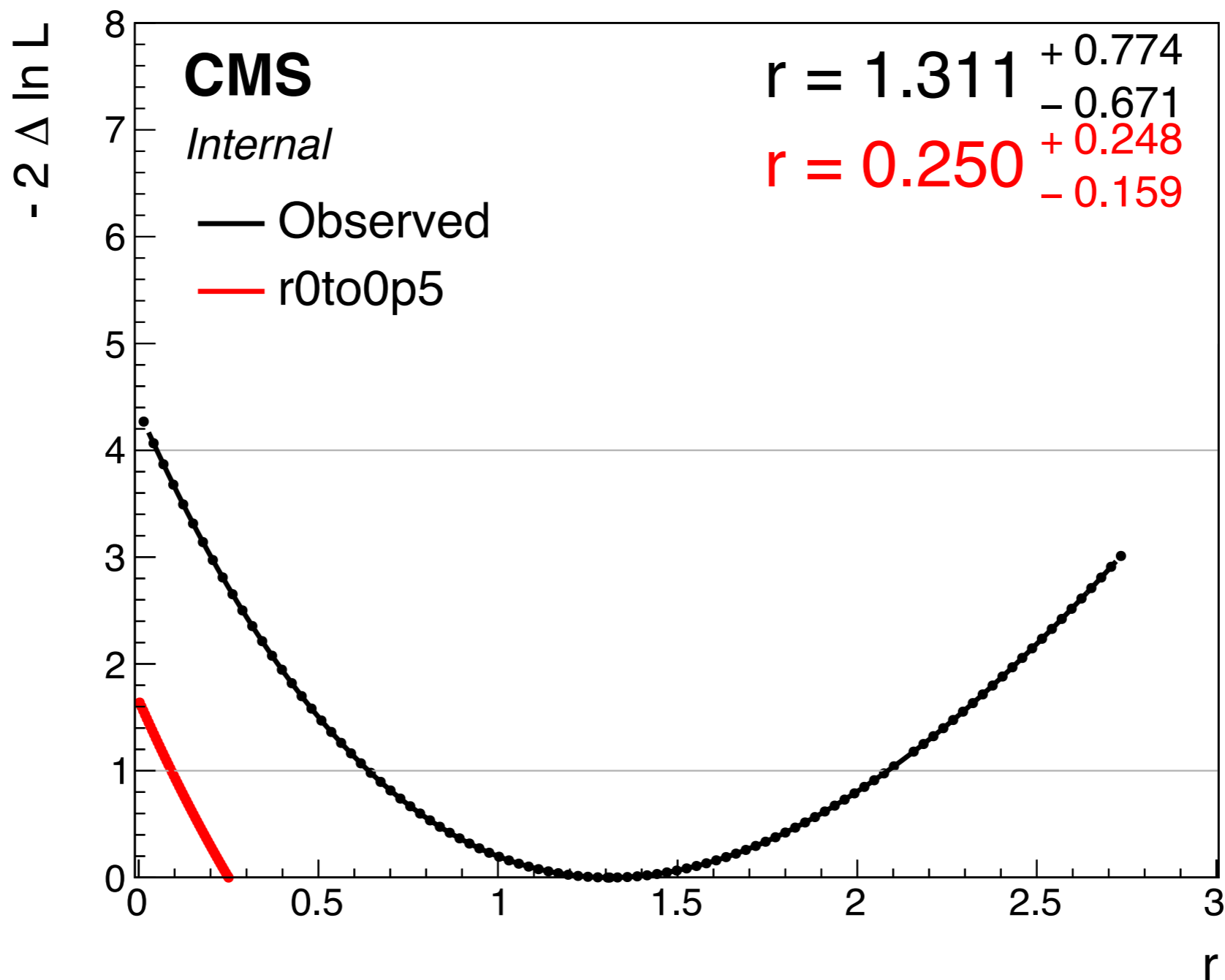
Plot overlaid on full range:

```
python ../../CombineHarvester/CombineTools/scripts/plot1DScan.py \
higgsCombineincl.MultiDimFit.mH125.root \
--others higgsCombiner0to0p5.MultiDimFit.mH125.root:"r0to0p5":2 \
--output "scan_overlay"
```

What went wrong?

What went wrong?

Two things: 1) initial value of r not in range, and 2) maximum likelihood (best fit) for subrange does not match full range

Stitching Likelihood Scans

Run initial fit and save snapshot:

```
combine -M MultiDimFit --mass 125 htt_tt_all.root --saveWorkspace -n Htt
```

This produces a file higgsCombineHtt.MultiDimFit.mH125.root with a workspace w containing the initial fit

Do scans, using snapshot for initial fit:

```
combine -M MultiDimFit --mass 125 -n r0to0p5 \
        --algo grid --points 100 --setParameterRange r=0,0.5 \
        -d higgsCombineHtt.MultiDimFit.mH125.root \
        -w w --snapshotName "MultiDimFit" --skipInitialFit
```

```
combine -M MultiDimFit --mass 125 -n r0p5to3p0 \
        --algo grid --points 100 --setParameterRange r=0.5,3.0 \
        -d higgsCombineHtt.MultiDimFit.mH125.root \
        -w w --snapshotName "MultiDimFit" --skipInitialFit
```

Plotting Scans

```
python ../../CombineHarvester/CombineTools/scripts/plot1DScan.py \
higgsCombineincl.MultiDimFit.mH125.root \
--others higgsCombiner0to0p5.MultiDimFit.mH125.root:"r0to0p5":2 \
        higgsCombiner0p5to3p0.MultiDimFit.mH125.root:"r0p5to3p0":4 \
--output "scan_overlay"
```
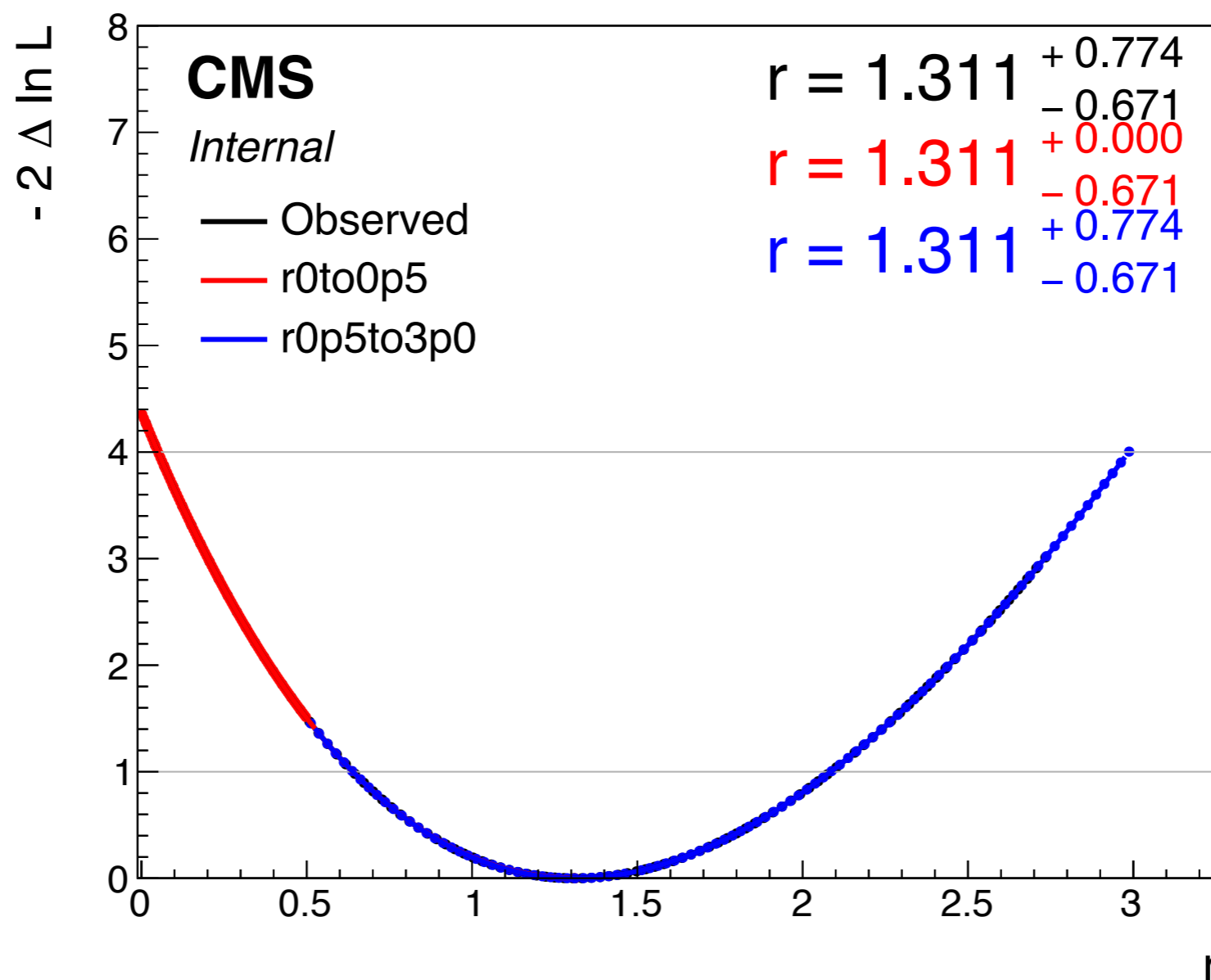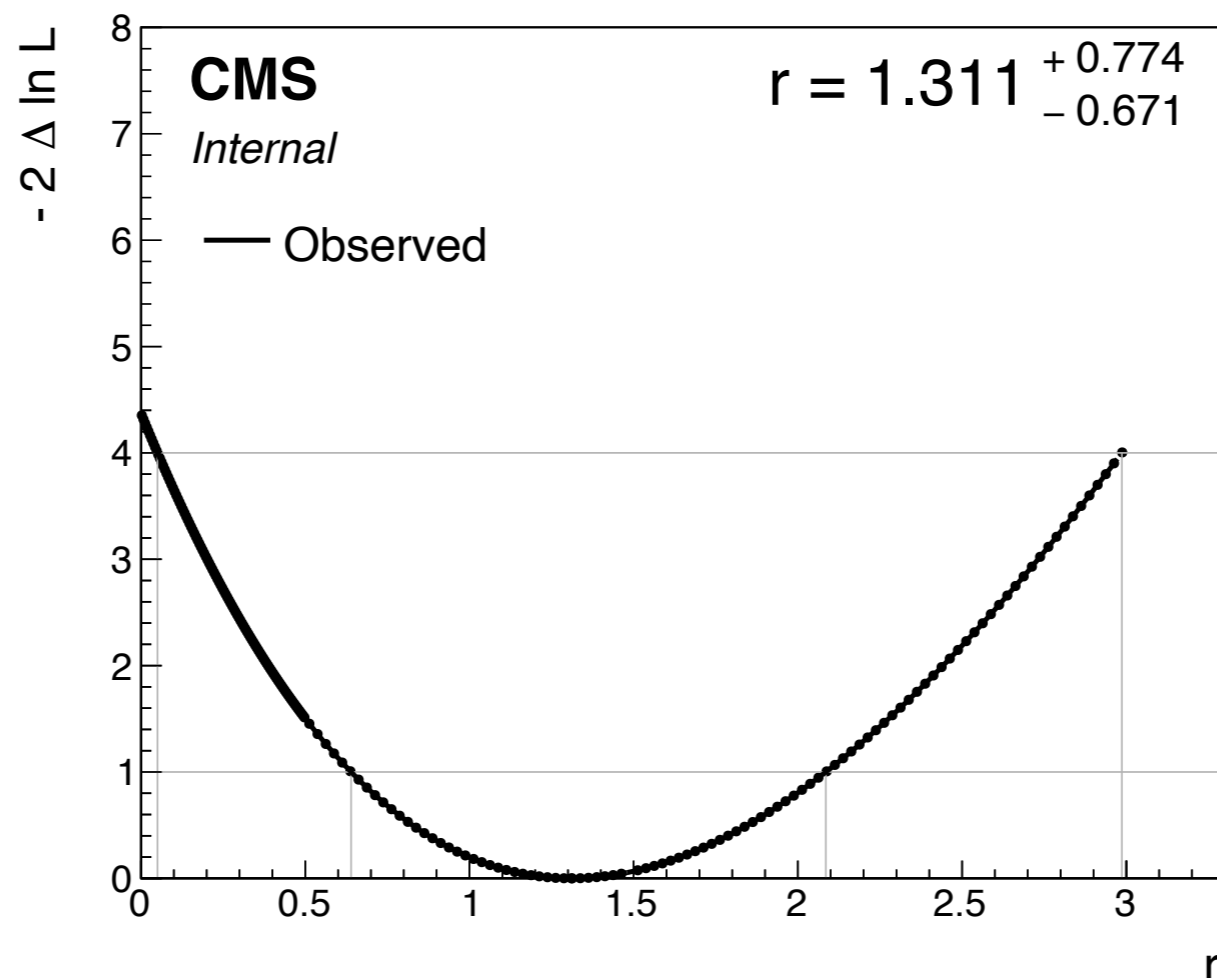
Invisible Stitching

Hadd outputs of (0,0.5) and (0.5,3.0) scans — can plot hadded tree as usual

```
hadd higgsCombinemerge.MultiDimFit.mH125.root \
higgsCombiner0to0p5.MultiDimFit.mH125.root \
higgsCombiner0p5to3p0.MultiDimFit.mH125.root

python ../../CombineHarvester/CombineTools/scripts/plot1DScan.py \
higgsCombinemerge.MultiDimFit.mH125.root
```

Extra: Nuisance Parameter Likelihood Scans

```
combineTool.py -M FastScan -w htt_tt_incl.root:w
```

- Produces a file nll.pdf with fast likelihood scans w.r.t. all parameters

- These fast scans use the conditional likelihood (fixed parameters, specifically fixing the non-scanned parameters to best fit value)

  - Intervals are not true confidence intervals

- Inspect output — does everything look ok?

  - What would it look like if thinks **weren't** ok?