Background Estimation Using Fake Rate Method(Part 3 Systematic Study)

Search for Four top in Tau Final States

Anshul Kapoor[1] Huiling Hua[1] Hongbo Liao[1] Xiaonan Hou[1]

[1]IHEP

November 4, 2022

Outline

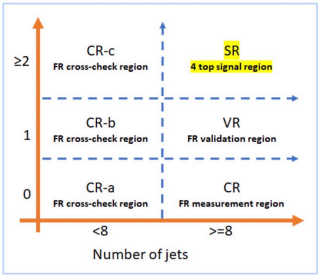
Section 1

Introduction

Systematic uncertainty from b jet multiplicity

- The only difference with SR(>=2 b jets) and CR(0 b jets) is the number of b jets
  - I feel like the probability of a event has more than 2 b jets doesn't impact the probability of the events's FR probability, since we are doing jet object removal with fakeable tau
- How to do investigate this?
  - Measure the FR in CRa and CRb
  - The difference is going to be our systematic uncertainty for b jet difference!
  - Check jet multiplicity impact with CR and CRa
  - We maybe should include CRc as validation region for FR too!

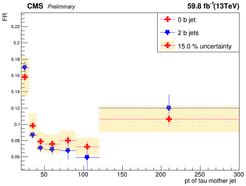
	N_τ	N_l	N_{jets}	N_b
SR	1	0	>=8	>=2
CR	1	0	>=8	0
VR	1	0	>=8	1
CRa	1	0	$< 8 \cap >= 6$	0
CRb	1	0	$< 8 \cap >= 6$	1
CRc	1	0	$< 8 \cap >= 6$	>=2

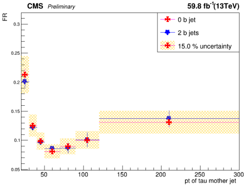


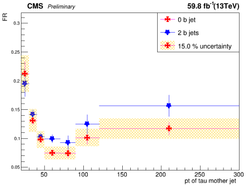
Section 2

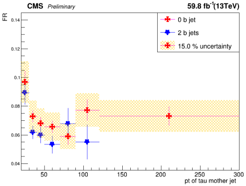
2018

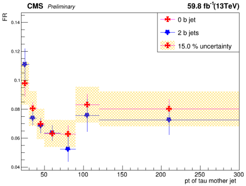
B jet multiplicity impact on FR(0 Vs >=2 b jets)

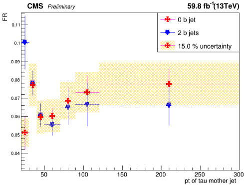






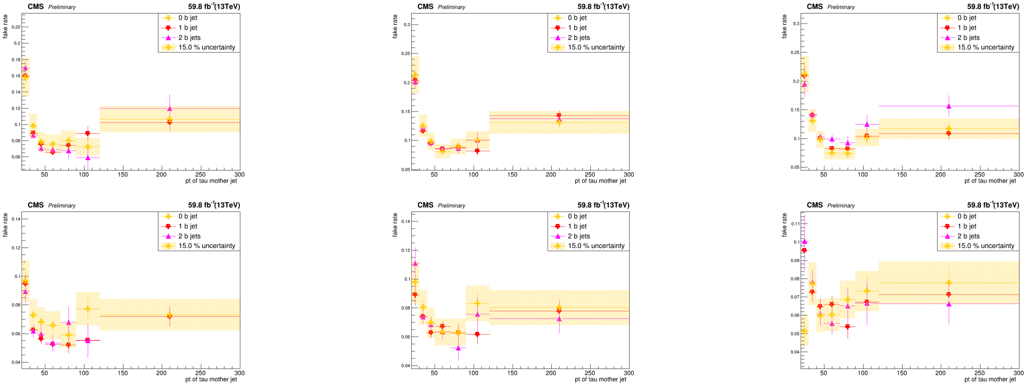






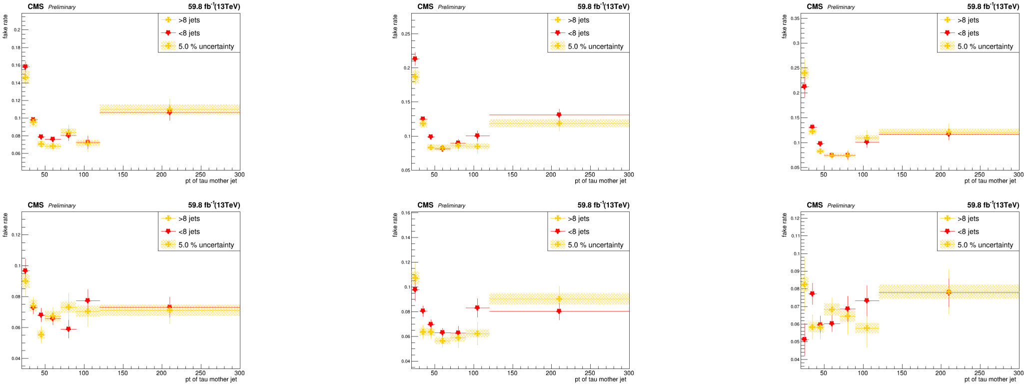
- From left to right: eta[0, 0.8]; eta[0.8, 1.6]; eta[1.6, 2.4]; from up to down: 1 prong, 2/3 prongs
- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- Measurement is really a tradeoff between statistic and systematic uncertainty! Found the best balance to achieve best accuracy

B jet multiplicity impact on FR(0 Vs 1 Vs >=2 b jets)



- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- We may want to further optimize the eta bin of FR measurement to reduce statistic uncertainty yet not introduce further inaccuracy

Jet multiplicity impact on FR(>8 Vs <8 jets)

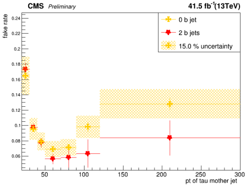


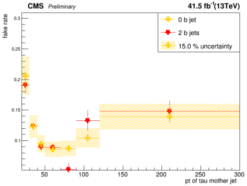
- Compare CR and CRa
- Jet multiplicity has very little impact on FR measurement from bins where statistics are enough
- We can safely assign the FR uncertainty measured in CRa and CRb to SR

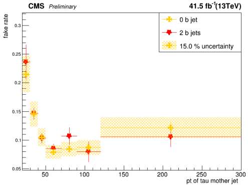
Section 3

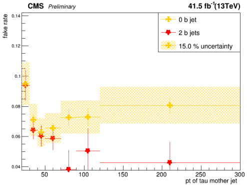
2017

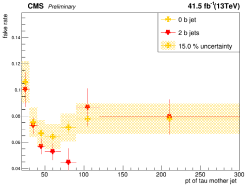
B jet multiplicity impact on FR(0 Vs >=2 b jets)

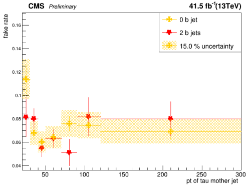






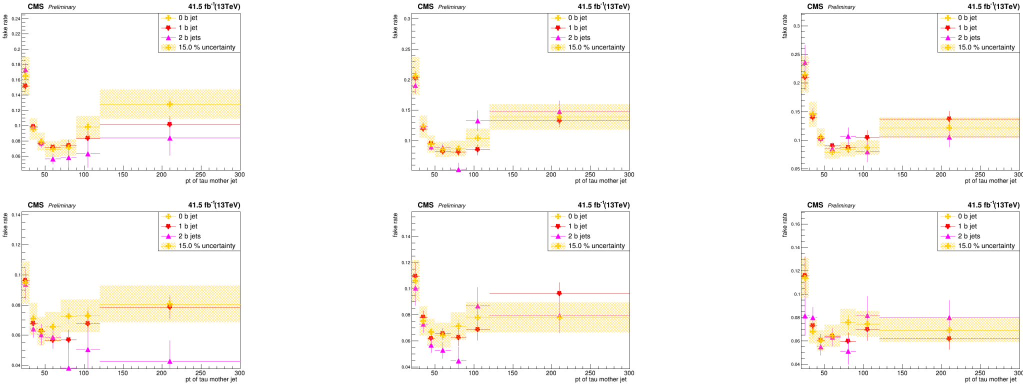






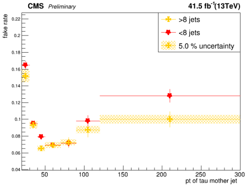
- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- Measurement is really a tradeoff between statistic and systematic uncertainty! Found the best balance to achieve best accuracy

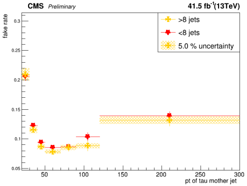
B jet multiplicity impact on FR(0 Vs 1 Vs >=2 b jets)

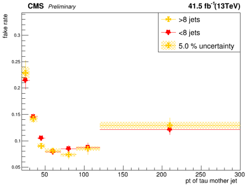


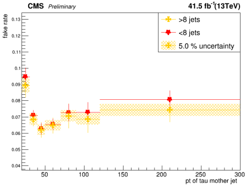
- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- We may want to further optimize the eta bin of FR measurement to reduce statistic uncertainty yet not introduce further inaccuracy

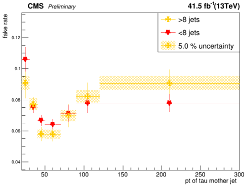
Jet multiplicity impact on FR(>8 Vs <8 jets)

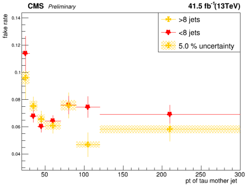










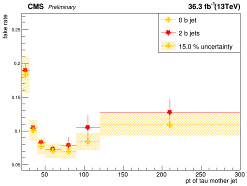


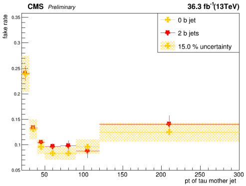
- Compare CR and CRa
- Jet multiplicity has very little impact on FR measurement from bins where statistics are enough
- We can safely assign the FR uncertainty measured in CRa and CRb to SR

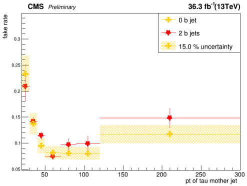
Section 4

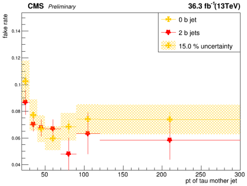
2016

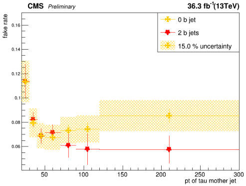
B jet multiplicity impact on FR(0 Vs >=2 b jets)

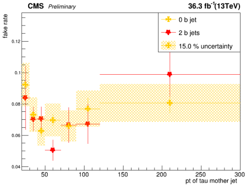






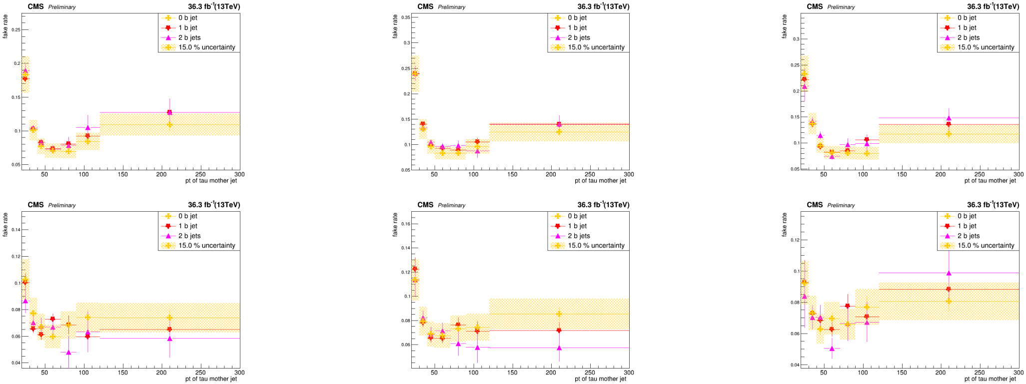






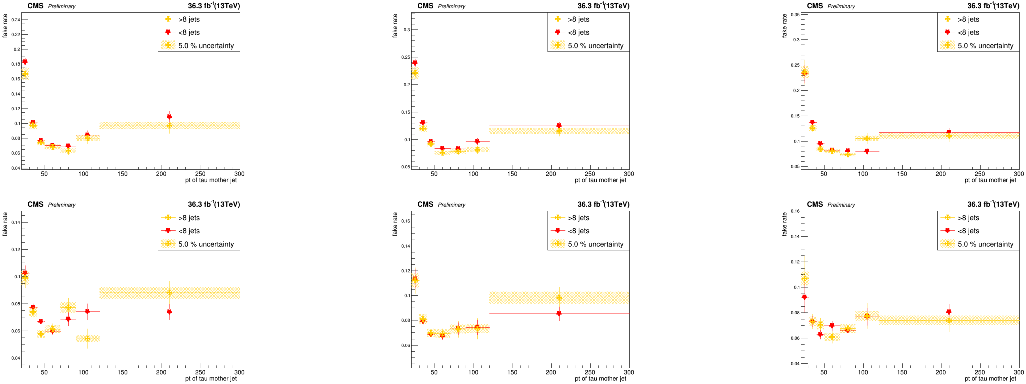
- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- Measurement is really a tradeoff between statistic and systematic uncertainty! Found the best balance to achieve best accuracy

B jet multiplicity impact on FR(0 Vs 1 Vs >=2 b jets)



- 15% uncertainty is more than enough
- For bins with low statistic, the under coverage is understandable due to the under coverage from the FR error
- We may want to further optimize the eta bin of FR measurement to reduce statistic uncertainty yet not introduce further inaccuracy

Jet multiplicity impact on FR(>8 Vs <8 jets)



- Compare CR and CRa
- Jet multiplicity has very little impact on FR measurement from bins where statistics are enough
- We can safely assign the FR uncertainty measured in CRa and CRb to SR

Section 5

Summary

Summary

- For all there years, the b jet multiplicity and jet multiplicity impact on FR are consistent
- Jet multiplicity has almost no impact on FR which makes sense, which means we can safely use the 15%(or even less) uncertainty assosiated with b jet in CR and SR
- B jet multiplicity has low impact on FR, we can assign 15% or 10%
- Other systematic uncertainty source to consider later
  - The substraction of MC gen tau events in SR could be a source of uncertainty
  - For now we are assumming all fakable tau are jets, we could have electrons or muons too.
  - How ever, the over assigned systematic uncertainty from b jet multiplicity may over this too. If we do validation in CRa and VR, everything covers fine we may not need to further investigate these
- We might want to optimize FR measurement with eta, lesser bins to reduce statistic uncertainty

Section 6

Back up

Event yield in FR cross check regions(2018)

regions	tttt	uncert	tt	uncert	fakeTau	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	s+b	data	data/bg	sensitivity
1tau0lSR	9.84	0.06	4103.11	13.80	6497.52	27.12	151.31	5.15	0.04	0.03	95.91	5.11	49.05	2.94	10906.78	-1.00	-0.00	0.09
1tau0lCR	0.04	0.00	97.49	2.14	7602.87	28.89	4.82	1.05	0.02	0.02	1.38	0.61	98.88	3.29	7805.50	7744.00	0.99	0.00
1tau0lVR	0.58	0.01	1476.86	8.28	10542.25	34.27	43.62	3.06	0.08	0.04	44.03	3.47	80.07	3.34	12187.50	12124.00	0.99	0.01
1tau0lCRa	0.02	0.00	184.69	2.93	14207.78	39.10	6.39	1.12	0.24	0.07	7.13	1.40	275.49	5.79	14681.72	15539.00	1.06	0.00
1tau0lCRb	0.28	0.01	2854.48	11.52	17904.16	44.24	75.52	4.13	0.12	0.05	123.07	5.79	226.36	6.73	21184.00	21637.00	1.02	0.00
1tau0lCRc	2.65	0.03	7396.35	17.95	11647.42	36.30	224.28	6.48	0.18	0.05	241.70	7.77	123.99	4.41	19636.56	19529.00	0.99	0.02

Table 1: 1tau0l verison 4, fake tau from FR, other MC requiring gen tau

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	total MC	data	data/MC
1tau0lSR	12.97	0.06	7601.83	20.46	3354.92	418.08	254.90	7.10	0.06	0.03	170.97	6.83	53.85	3.05	11449.50	-1.00	-0.00
1tau0lCR	0.05	0.00	218.73	3.55	4751.14	546.74	8.57	1.32	0.07	0.04	4.39	1.09	110.04	3.40	5093.00	7744.00	1.52
1tau0lVR	0.81	0.02	2927.97	12.81	5108.74	559.36	77.91	4.36	0.14	0.06	84.42	4.80	90.97	3.58	8290.97	12124.00	1.46
1tau0lCRa	0.02	0.00	375.67	4.59	9628.37	780.28	11.07	1.60	0.54	0.11	16.68	2.13	307.40	6.19	10339.75	15539.00	1.50
1tau0lCRb	0.35	0.01	5074.79	16.64	11996.31	829.07	124.54	5.69	0.30	0.08	217.09	7.69	250.10	7.01	17663.47	21637.00	1.22
1tau0lCRc	3.20	0.03	11312.71	24.59	6045.33	670.40	332.54	8.44	0.23	0.07	358.97	9.89	110.76	4.58	18163.74	19529.00	1.08

Table 2: 1tau0l verison 4, all MC prediction

- TTTT signal negligible in control regions
- CRc could also be a great validation region

Event yield in FR cross check regions(2017)

regions	tttt	uncert	tt	uncert	fakeTau	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	s+b	data	data/bg	sensitivity
1tau0lSR	6.62	0.04	2799.39	10.96	4560.70	23.22	102.49	3.76	0.06	0.03	68.21	4.16	34.70	2.44	7572.16	-1.00	-0.00	0.08
1tau0lCR	0.03	0.00	97.02	2.04	11530.80	36.51	3.87	0.82	0.07	0.03	2.49	0.79	109.73	3.70	11744.00	11672.00	0.99	0.00
1tau0lVR	0.43	0.01	1027.06	6.64	8504.51	31.33	35.95	2.40	0.03	0.02	27.77	2.66	67.20	3.15	9662.94	9463.00	0.98	0.00
1tau0lCRa	0.01	0.00	169.75	2.69	21387.56	49.42	5.03	0.92	0.29	0.07	7.52	1.38	287.88	6.59	21858.04	23144.00	1.06	0.00
1tau0lCRb	0.17	0.01	1833.03	8.88	13252.28	38.83	49.83	3.13	0.19	0.05	89.36	4.77	157.60	5.11	15382.45	15725.00	1.02	0.00
1tau0lCRc	1.71	0.02	4767.08	13.89	7647.56	30.00	147.61	4.80	0.07	0.02	162.62	6.17	82.28	3.43	12808.93	12484.00	0.97	0.02

Table 3: 1tau0l verison 4, fake tau from FR, other MC requiring gen tau

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	S+B	data	data/MC
1tau0lSR	8.69	0.05	5131.36	16.33	2321.60	383.07	169.57	5.23	0.09	0.04	119.12	5.50	39.37	2.64	7789.80	-1.00	-0.00
1tau0lCR	0.04	0.00	202.16	3.29	4833.71	432.62	7.04	1.22	0.12	0.04	4.77	1.10	123.55	3.94	5171.38	11672.00	2.26
1tau0lVR	0.58	0.01	2028.57	10.36	4077.31	529.73	62.90	3.41	0.13	0.04	55.90	3.78	75.80	3.38	6301.18	9463.00	1.50
1tau0lCRa	0.02	0.00	315.27	4.05	11910.91	956.39	8.30	1.23	0.60	0.10	15.29	1.97	319.51	6.93	12569.90	23144.00	1.84
1tau0lCRb	0.22	0.01	3212.17	12.83	7443.91	665.52	83.36	4.12	0.28	0.07	149.21	6.17	178.98	5.47	11068.14	15725.00	1.42
1tau0lCRc	2.06	0.02	7231.29	19.09	3612.99	428.35	202.70	6.08	0.16	0.05	238.43	7.81	73.28	3.66	11360.90	12484.00	1.10

Table 4: 1tau0l verison 4, all MC prediction

- TTTT signal negligible in control regions
- CRc could also be a great validation region

Event yield in FR cross check regions(2016)

regions	tttt	uncert	tt	uncert	fakeTau	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	s+b	data	data/MC	sensitivity
1tau0lSR	5.22	0.04	1996.31	9.71	5310.96	25.17	69.19	3.27	0.02	0.01	48.67	3.58	26.91	2.25	7457.28	-1.00	-0.00	0.06
1tau0lCR	0.05	0.00	145.11	2.62	15042.49	41.86	5.53	0.86	0.09	0.03	6.11	1.27	138.26	4.67	15337.65	15300.00	1.00	0.00
1tau0lVR	0.50	0.01	959.62	6.73	11456.29	36.62	31.92	2.45	0.04	0.02	25.87	2.62	61.36	3.21	12535.58	12847.00	1.02	0.00
1tau0lCRa	0.03	0.00	282.24	3.65	28580.63	57.37	9.51	1.34	0.24	0.04	12.38	1.81	455.92	9.41	29340.95	30807.00	1.05	0.00
1tau0lCRb	0.22	0.01	1976.47	9.68	18878.89	46.98	46.93	3.25	0.15	0.03	76.03	4.48	174.40	5.56	21153.09	22722.00	1.07	0.00
1tau0lCRc	1.48	0.02	4002.54	13.22	8886.77	32.63	115.58	4.36	0.08	0.02	121.15	5.33	72.57	2.88	13200.17	13577.00	1.03	0.01

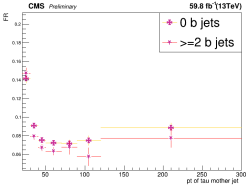
Table 5: 1tau0l verison 4, fake tau from FR, other MC requiring gen tau

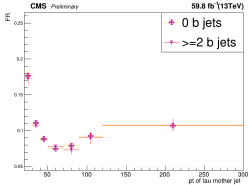
regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	S+B	data	data/MC
1tau0lSR	6.96	0.04	3798.06	14.45	2167.32	432.00	127.35	4.64	0.06	0.02	89.33	4.86	32.65	2.49	6221.73	-1.00	-0.00
1tau0lCR	0.07	0.00	312.25	4.18	7198.95	530.76	7.94	1.29	0.19	0.04	11.89	1.77	161.83	5.17	7693.10	15300.00	1.99
1tau0lVR	0.69	0.01	1931.27	10.34	5663.20	742.55	56.38	3.48	0.11	0.03	51.41	3.69	73.19	3.61	7776.25	12847.00	1.65
1tau0lCRa	0.03	0.00	545.98	5.47	23632.82	1535.84	15.73	1.88	0.68	0.07	23.56	2.49	520.92	10.02	24739.74	30807.00	1.25
1tau0lCRb	0.27	0.01	3536.78	13.85	10408.22	1112.40	76.78	4.38	0.35	0.05	147.69	6.23	201.26	6.06	14371.34	22722.00	1.58
1tau0lCRc	1.80	0.02	6127.01	18.13	4564.54	826.47	167.88	5.80	0.12	0.03	182.96	6.95	59.57	3.09	11103.87	13577.00	1.22

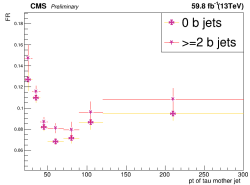
Table 6: 1tau0l verison 4, all MC prediction

- TTTT signal negligible in control regions
- CRc could also be a great validation region

FR in CR1 and CR2, no prong division







Statistic uncertainty of FR

- $FR = \frac{N^{notGen\tau}_{VVL\tau} = N^{data}_{VVL\tau} - N^{true\tau}_{VVL\tau}(MC)}{N^{notGen\tau}_{T\tau} = N^{data}_{T\tau} - N^{true\tau}_{T\tau}(MC)}$
- Tricky here because we are dividing two poisson distribution
- Error propagation case(the default calculated by root Divide method)
  - The error of each bin has to be small compared to mean(essential assumption for error propagation)
  - denominator and numerator uncorrelated(no correlation matrix in equation)
  - $\sigma^2(FR) = \frac{1}{y^2} * \sigma^2(X) + \frac{x^2}{y^4} * \sigma^2(Y)$ (X is numerator, Y is denominator)
- When event count it low we have other methods available suggested by CMS statistic group
- For simplicity use default provided by ROOT Divide
- When we substrat MC for