Background Estimation Using Fake Rate Method(New Lepton Definition)

Search for Four top in Tau Final States

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

[1]IHEP

February 15, 2023

Outline

1. Introduction
2. Fake rate measured in CR
3. Application and Validation
4. Summary

Section 1

Introduction

Update from previous FR study

- Switch lepton definition to SS lepton definition(Top lepton MVA used)
- veto SS tight lepton in 1tau0l
- Eta binning for FR measurement became [0, 1.5], [1.5, 2.4]

Event yield comparision

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	bg	sensitivity	data	data/MC
1tau0ISR	7.013	0.044	3687.118	14.237	1654.362	186.398	124.802	4.574	0.050	0.018	85.974	4.767	30.533	2.394	5582.838	0.094	-1.000	-0.000
1tau0ICR	0.070	0.005	297.755	4.089	6911.947	527.270	8.050	1.268	0.159	0.032	9.979	1.620	148.107	5.003	7375.997	0.001	15300.000	2.074

Table 1: old tight lepton veto

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	total bg	sensitivity	data	data/MC
1tau0ISR	6.924	0.044	3626.183	14.134	1653.891	186.398	123.399	4.523	0.055	0.019	83.523	4.698	30.317	2.426	5517.368	0.093	-1.000	-0.000
1tau0ICR	0.067	0.005	295.333	4.078	6864.155	525.777	8.105	1.292	0.159	0.032	9.439	1.575	144.514	4.964	7321.705	0.001	15265.000	2.085

Table 2: new tight lepton veto

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	bg	sensitivity	data	data/MC
1tau0ISR	4.128	0.034	2651.642	12.116	1256.072	165.186	84.861	3.844	0.020	0.012	63.789	4.107	20.444	1.855	4076.828	0.065	-1.000	-0.000
1tau0ICR	0.041	0.004	217.819	3.511	5995.912	512.402	5.375	1.026	0.119	0.028	7.614	1.415	119.418	4.466	6346.258	0.001	13229.000	2.085

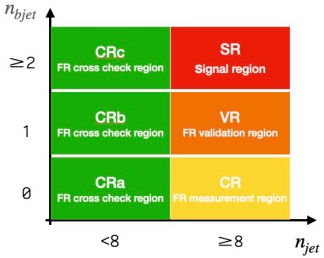
Table 3: new loose lepton veto

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	bg	sensitivity	data	data/MC
1tau0ISR	4.338	0.035	2801.903	12.442	1327.875	169.919	89.807	3.879	0.027	0.014	64.612	4.134	23.712	2.108	4307.937	0.066	-1.000	-0.000
1tau0ICR	0.050	0.004	238.851	3.675	6441.435	518.785	5.851	1.119	0.154	0.032	7.910	1.445	133.855	4.761	6828.056	0.001	14235.000	2.085

Section 2

Fake rate measured in CR

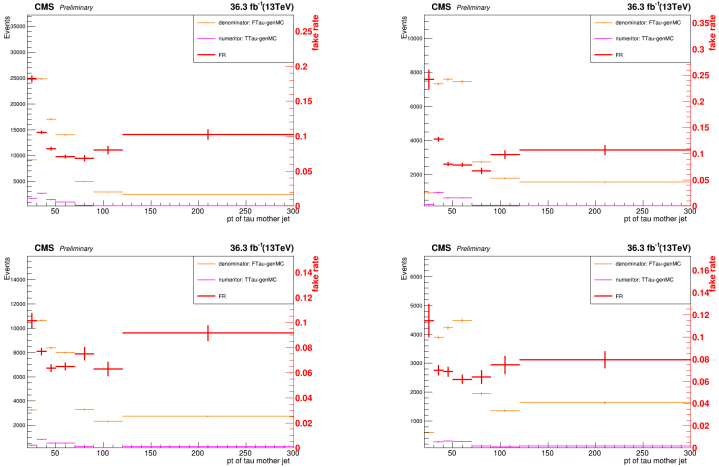
Calculate fake rate in CR(2016)



- $FR = \frac{N_{T\tau}^{fake\tau}}{N_{VVL\tau}^{\tau}} = \frac{N_{T\tau}^{data} - N_{T\tau}^{true\tau}(MC)}{N_{VVL\tau}^{data} - N_{VVL\tau}^{true\tau}(MC)}$
  - $N_{VVL\tau}^{\tau} = N_{VVL\tau}^{data} - N_{VVL\tau}^{true\tau}(MC)$
    - MC: all bg except qcd, assign 20% uncertainty for this
    - VVL tau here is our fakable tau
  - $N_{T\tau}^{fake\tau} = N_{T\tau}^{data} - N_{T\tau}^{true\tau}(MC)$

FR binning into tauJet pt, tauJet eta, tau prong

- 2 prongs is actually 3 prongs with one track lost during reconstruction
- Pt binning: [20.0, 30, 40.0, 50, 70.0, 90.0, 120.0, 300.0]
- From left to right in eta regions: [0, 1.5], (1.5, 2.4]
- From top to bottom: 1 prong, 2/3 prongs



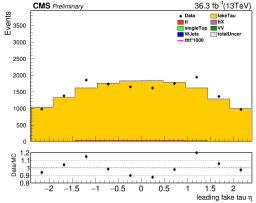
Section 3

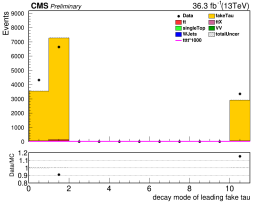
Application and Validation

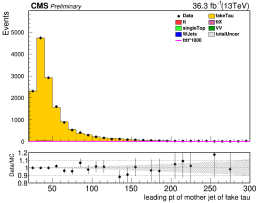
Apply fake rate in application region

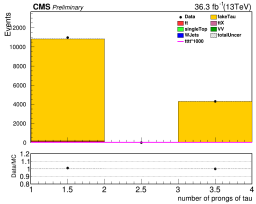
- Use the fake rate we get from CR and apply here
- $N_{fake\tau} = N^{notTrue\tau}_{VVL\tau} \times FR = N^{notGen\tau}_{VVL\overline{T}} \times \frac{FR}{1-FR} = (N^{data}_{VVL\overline{T}} - N^{gen\tau(MC)}_{VVL\overline{T}}) \times \frac{FR}{1-FR}$
- FR can be used to predict shape too
  - FR can be interpreted as probability of a fake tau passing the tight tau selection,
  - FR/1-FR as weight to each event(data-genMC) in VVLNotT region
  - Critical in code implementation to ensure that pt is within the range of FR measurement

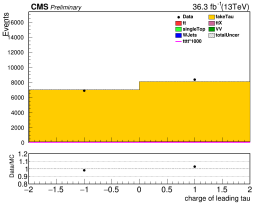
Closure test in CR: tau varaibles

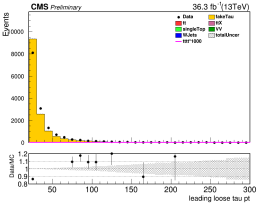






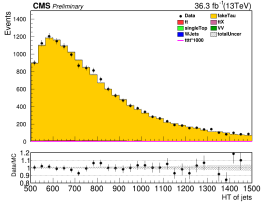


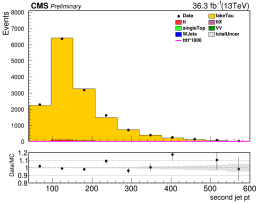


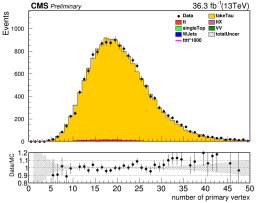


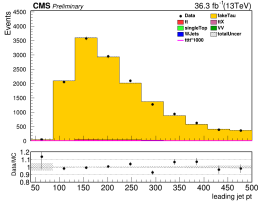
- FR dependent on tau decay mode too
- Only statistic uncertainty, no uncertainty from FR measurement considered

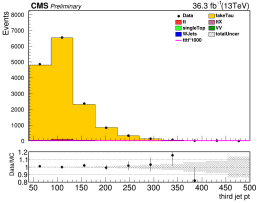
Closure test in CR: jet varaibles

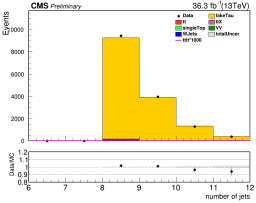




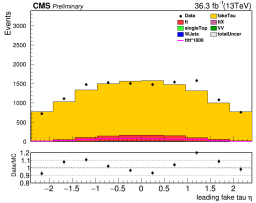


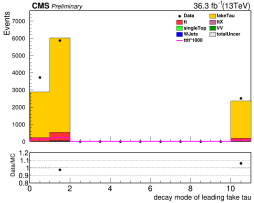


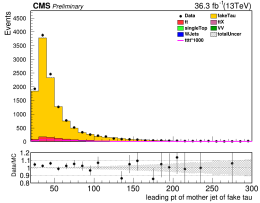


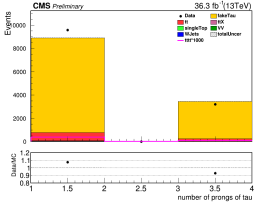


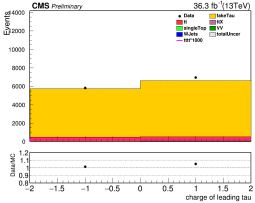
Closure test in VR: tau varaibles

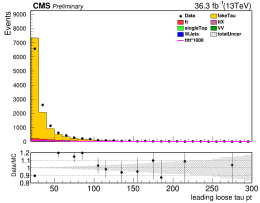






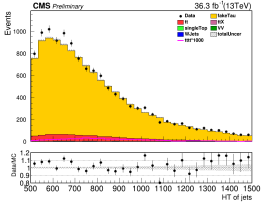


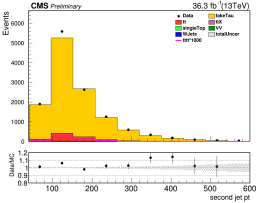


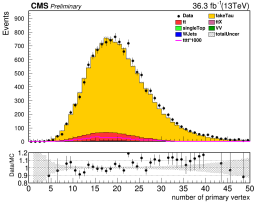


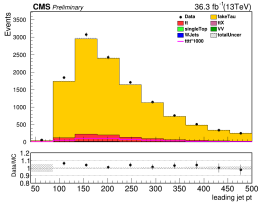
- FR dependent on tau decay mode too
- Only statistic uncertainty, no uncertainty from FR measurement considered

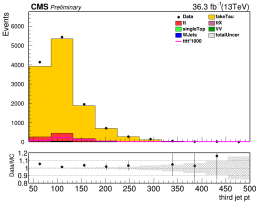
Closure test in VR: jet varaibles

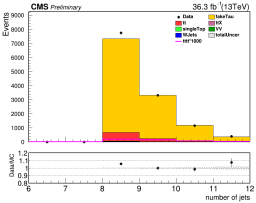




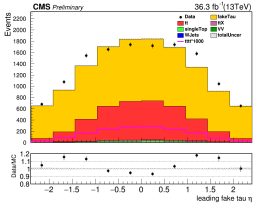


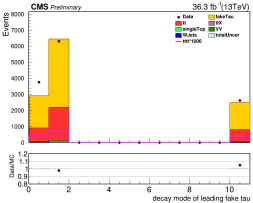


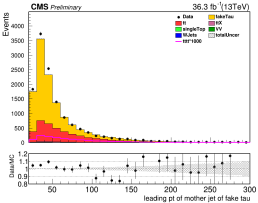


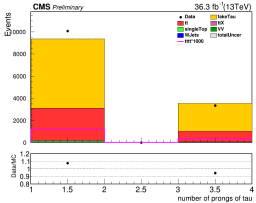


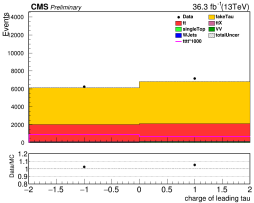
Closure test in CRc: tau varaibles

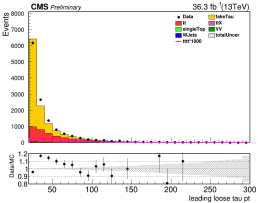






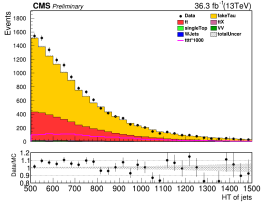


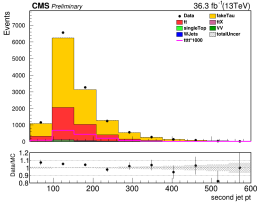


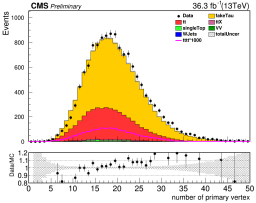


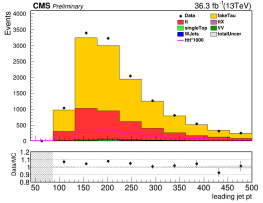
- FR dependent on tau decay mode too

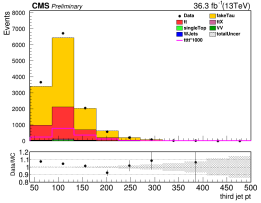
Closure test in CRc: jet varaibles

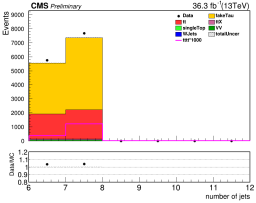




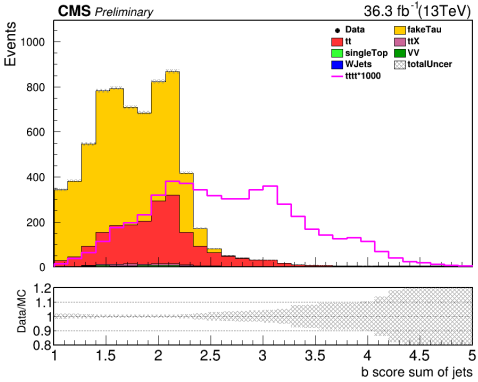








Discriminant in SR



Tasks and Progress

- Done
  - Study of lepton veto in 1tau0l
  - Repeat FR study in 2016
- To do
  - Apply all MC correction(1 week)
  - Repeat 1tau1l BDT study (2 days)
  - Systematic study (1 week)
- Conclusion
  - A little bit below the progress, need to move faster with later tasks

Questions

- How to measure the selection SF on veto objects?

Section 4

Summary

Section 5

Back up

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

Statistic uncertainty of fakable tau yield(applied bin by bin)

- $N_{fake\tau} = (N^{data}_{VVL\overline{T}} - N^{gen\tau(MC)}_{VVL\overline{T}}) \times \frac{FR}{1-FR}$
- According to error propagation
  - $\frac{dN}{dFR} = n * \frac{1}{(1-FR)^2}$ $(n = N^{data}_{VVL\overline{T}} - N^{gen\tau(MC)}_{VVL\overline{T}})$
  - $\sigma^2(N) = \frac{n^2}{(1-FR)^4} * \sigma^2(FR) + \frac{FR^2}{(1-FR)^2} * \sigma^2(n)$
- For the sum, simply sum the σ^2 (error propagation)
- Error propagation suitable for large number of events, might need better method for bins with low statistic

Uncertainty for estimated fake tau shape(applied as event weight)

- fake rate weight for data: $W_{FR} = \frac{FR}{1-FR}$
- $\sigma^2(W_{FR}) = (\frac{dW}{dFR})^2 \times \sigma^2(FR) = \frac{1}{(1-FR)^4} \times \sigma^2(FR)$
- $W \pm \sigma(W) = \frac{FR}{1-FR} \pm \frac{\sigma(FR)}{(1-FR)^2}$
- Vary the fake rate event weight in the above way to get the up and down shape of fake tau caused by FR measurement
- For now the $\sigma(FR)$ only contains statistic uncertainty of the FR measurement, systematic uncertainty to be added

Uncertainty of data/MC

- $\sigma^2_{total} = \sigma^2_{statistic} + \sigma^2_{systematic}$
- $\sigma^2_{systematicUp} = \sum_{i=1}(n_{iWeight_up} - n_{nominal})^2$
  - i is the i source of uncertainty, for example FR uncertainty for fake tau
- Fake tau
  - Other systematic uncertainty like pileup reweighting and prefiring reweighting does not impact the shape of fake tau since it's estimated from data
  - only the FR uncertainty impact the fake tau
  - The systematic uncertainty of fake tau: only from FR measurement
- Other bgs except fake tau
  - For now: prefiring reweighting, pileup reweighting
  - Object selection corrections:

Tau decay

- 1 prong(charge hardrons): 47%
- 3 prong: 15 %
- leptonic: 35%

Fake rate in CR

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	total MC	data	data/MC
1tau0lSR	6.924	0.044	3626.183	14.134	1653.891	186.398	123.399	4.523	0.055	0.019	83.523	4.698	30.317	2.426	5524.291	-1.000	-0.000
1tau0lCR	0.067	0.005	295.333	4.078	6864.155	525.777	8.105	1.292	0.159	0.032	9.439	1.575	144.514	4.964	7321.772	15265.000	2.085
1tau0lCRLTau	0.283	0.009	1747.787	10.470	87486.106	2743.190	44.302	3.205	0.877	0.079	47.552	3.541	271.058	6.855	89597.966	163650.000	1.826
1tau0lVR	0.689	0.014	1841.239	10.115	52195.39	732.227	50.619	3.362	0.137	0.033	48.720	3.587	63.893	3.251	7224.836	12800.000	1.772
1tau0lVRLTau	2.448	0.027	9816.093	24.747	46779.934	2175.477	226.647	7.496	0.500	0.060	244.805	8.046	123.230	4.542	57193.658	126266.000	2.208

Table 5: 1tau0l

regions	tttt	uncert	tt	uncert	qcd	uncert	ttX	uncert	VV	uncert	singleTop	uncert	WJets	uncert	total MC
1tau0lSRGen	5.271	0.038	1929.855	9.552	0.000	0.000	68.877	3.197	0.028	0.013	45.845	3.472	25.926	2.246	2075.802
1tau0lCRGen	0.048	0.004	138.473	2.562	7.063	4.721	5.033	0.790	0.067	0.021	4.742	1.119	125.801	4.521	281.227
1tau0lCRLTauGen	0.085	0.005	238.277	3.383	131.724	56.934	8.904	1.120	0.088	0.024	6.320	1.289	203.473	5.864	588.871
1tau0lVRGen	0.500	0.012	923.963	6.606	0.171	0.171	29.461	2.340	0.056	0.023	24.276	2.536	54.702	2.916	1033.128
1tau0lVRLTauGen	0.838	0.016	1540.776	8.566	98.259	43.219	47.973	3.142	0.076	0.025	37.633	3.153	88.316	3.763	1813.870

Table 6: Require the tau to be true hadronic tau too