Estimation of the QCD background yield in the 1τ 0L category

Huiling Hua [1] **Fabio Iemmi** [1]
Hongbo Liao [1] Hideki Okawa [2] Yu Zhang [2]

[1]Institute of High Energy Physics (IHEP), Beijing

[2]Fudan University, Shanghai

August 18, 2021

QCD-enriched control region (CR)

- QCD is only dominant in 1τ 0L category
  - $\approx$ 50% of the background yield in 1τ 0L
- All the remaning **major backgrounds** ($t\bar{t}$ and $t\bar{t}$+X) **and signal involve** top quarks, i.e., **bottom quarks** in the final state
- **Revert the request on the number of b tagged jets in the event**

	N_{τ_h}	N_ℓ	N_{jets}	N_{bjets}
1τ 0L	1	0	≥ 8	≥ 2
1τ 0L ctl	1	0	≥ 8	0

- $N_{bjets} = 0$ is meant to reject all the top-related processes

Yields

	$t\bar{t}t\bar{t}$	$t\bar{t}$	QCD	$t\bar{t}$+X
1τ 0L SR	10	6371	7461	192
1τ 0L CR	0	294	8087	8

- The CR is QCD-dominated
- QCD events form 96% of the events in CR
- The **large QCD simulated yield** that we get **in CR should come from fake taus**

Fake rate method

- Inspired by EXO-19-015
- Estimate the background completely from data by doing

$$N_{\text{fake-}\tau} = \sum_{p_T,\eta} N_{\text{fake-}\tau}(p_T,\eta) = \sum_{p_T,\eta} \left[N_{\text{F},\overline{\text{T}}}(p_T,\eta) \times \frac{\text{FR}(p_T,\eta)}{1-\text{FR}(p_T,\eta)} \right]$$

- $N_{\text{F},\overline{\text{T}}}(p_T,\eta)$ is computed in the application region (AR)
- $\text{FR}(p_T,\eta)$ is computed in the CR
- Parametrize as a function of p_T, η of fakeable tau
- Binning in (p_T,η): $p_T \in [20,30,75,150,300,\text{Inf}]$; $\eta \in [0,1.5,2.3]$

Closure test in MC QCD

- Compute FR in CR, **apply the method in** the same **CR**
- **Compare with number** of events in CR **you count from MC**
- These numbers should close

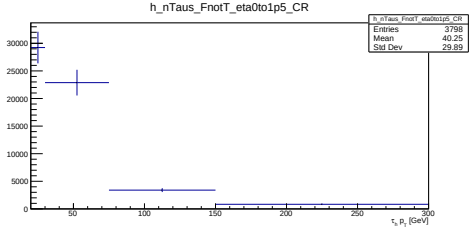
	Value	Raw entries
Counting	7979 ± 1350	547
Fake rate method	8636 ± 2321	–

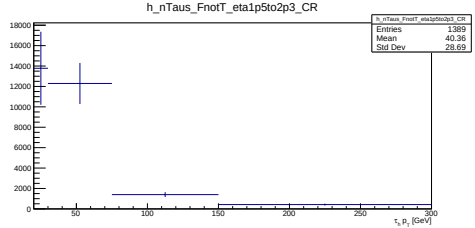
- Values are in **agreement** within the uncertainties [1], **closure is not perfect** (8% discrepancy)
  - Most likely **due to approximations** in weighting and summing `TEfficiency` objects
  - See my discussion with ROOT developer Lorenzo Moneta here

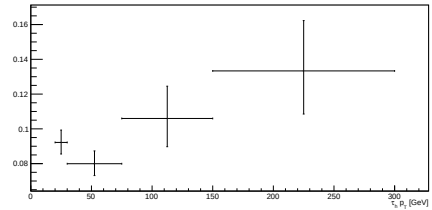
[1] See backup for error computation

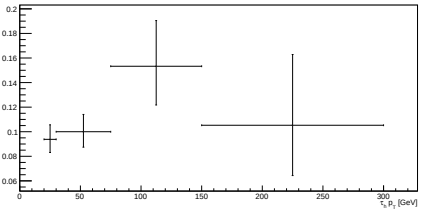
Closure test on MC QCD: plots

- Upper row: number of fakeable-not-tight taus, lower row: fake rates









Fake rate method in MC QCD

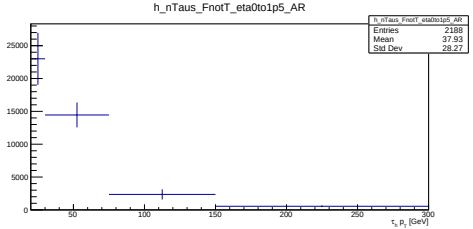
- Compute FR in CR, **apply the method in** the **application region** (same as signal region, but use fakeable-not-tight taus)
- **Compare with number** of events in SR **you count from MC**

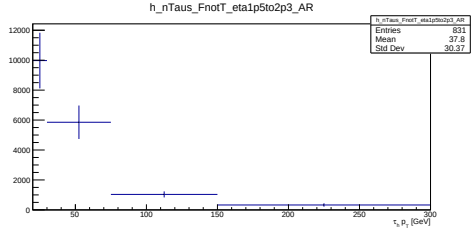
	Value	Raw entries
Counting	7461 ± 1681	315
Fake rate method	5887 ± 1782	–

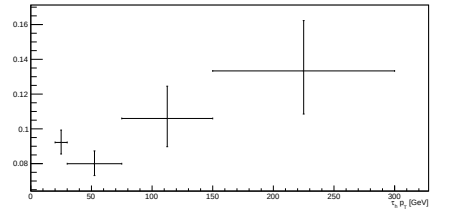
- Values are in **agreement within the uncertainties**
- Uncertainties are big due to poor statistics in MC samples

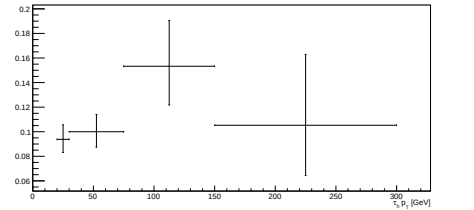
Fake rate method in MC QCD: plots

- Upper row: number of fakeable-not-tight taus, lower row: fake rates









Closure test in data

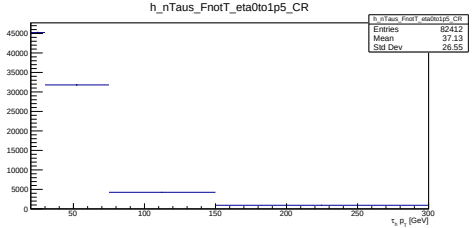
- Compute FR in CR, **apply the method in** the same **CR**
- **Compare with number** of events in CR **you count from data**
- This should close (at least approximately)

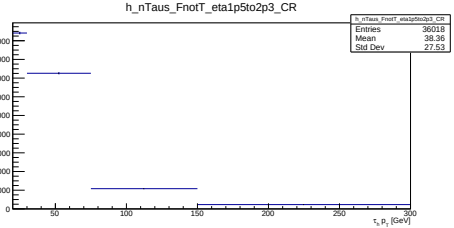
	Value	Raw entries
Counting	11561 ± 108	11561
Fake rate method	11561 ± 384	–

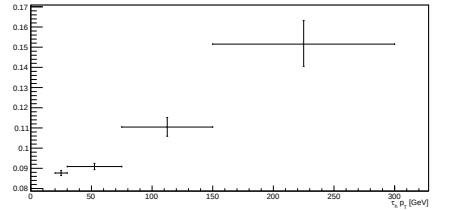
- Values are in **agreement** within the uncertainties, **perfect closure**
  - No weighting of any kind of objects is needed for data

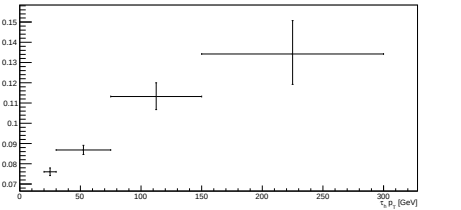
Closure test in DATA: plots

- Upper row: number of fakeable-not-tight taus, lower row: fake rates









Fake rate method in data

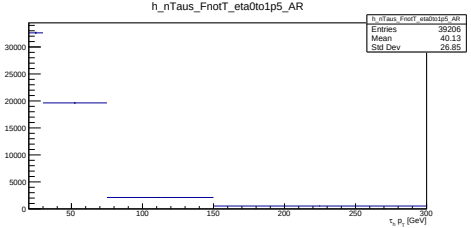
- Compute FR in CR, **apply the method in** the **application region** (same as signal region, but use fakeable-not-tight taus)
- **Do not compare with number** of events in SR **you count from data**: **we are blinded**
- **IMPORTANT!**: take care of subtracting $t\bar{t}$ and $t\bar{t}$+X from $N_{F,\overline{T}}(p_T, \eta)$

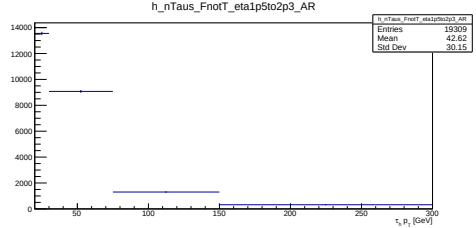
	Value	Raw entries
Counting	–	–
Fake rate method	7679 ± 273	–

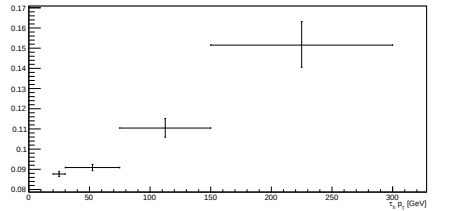
- The **estimated yield** from FR method is i**n agreement with QCD MC predictions**
- But it comes with **way lower uncertainty!** (4% vs 23%)

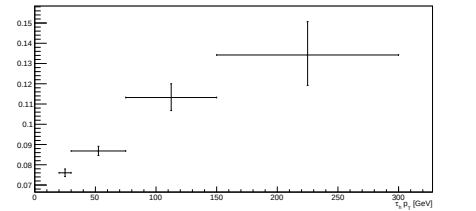
Fake rate method in DATA: plots

- Upper row: number of fakeable-not-tight taus, lower row: fake rates









Conclusions

- Developed a **method to estimate the QCD background** yield in 1τ 0L category **completely from data**
- The prediction of MC QCD and the value from FR method agree within uncertainties

	MC QCD	FR method
Yield	7461 ± 1681	7679 ± 273

- **Uncertainty** on the yield **is greatly reduced**
- Statistical uncertainty on MC QCD yield is 23%; customary to assign flat 50% uncertainty in datacards to cover big theoretical uncertainties
- Our method uses data only and results in **4% uncertainty** only

Backup slides

Error computation for MC QCD yields

- Simple counting of number of simulated QCD events
- Set up a histogram `TH1F * h_QCD_yield = new TH1F("h_QCD_yield", "h_QCD_yield", 1, 0, 3000);`
  - Just one bin in tau p_T
- Fill it with tau p_T weighting for the xsec of each QCD slice: `h_QCD_yield->Fill(mytausF->at(0).Pt(), LUMI2016*xsec/gen_sum_of_weights);`
- Get yield and corresponding error using `Integral(-1,-1)` and `GetBinError(1)` methods

Error computation for fake rate yields

- Things are a bit more convoluted for fake rate yields
- Yields come from the formula

$$N_{\text{fake-}\tau} = \sum_{p_T,\eta} N_{\text{fake-}\tau}(p_T, \eta) = \sum_{p_T,\eta} \left[N_{\text{F},\overline{\text{T}}}(p_T, \eta) \times \frac{\text{FR}(p_T, \eta)}{1 - \text{FR}(p_T, \eta)} \right]$$

- The correlation between each of the (p_T, η) bin contents is not trivial to judge: are they correlated or not? Similarly, are the FR correlated or not?
- Decide to use the simplest (and in worst case, over-conservative) error propagation formula
- Our formula is a sum of products
- For products, sum the relative uncertainties; for sums, sum the absolute uncertainties

Error computation for fake rate yields

- For each product

$$P(p_T, \eta) = N_{F,\overline{T}}(p_T, \eta) \times \frac{FR(p_T, \eta)}{1 - FR(p_T, \eta)}$$

the relative uncertainty is

$$\frac{\Delta P}{P} = \frac{\Delta N_{F,\overline{T}}}{N_{F,\overline{T}}} + \frac{\Delta FR}{FR} + \frac{\Delta FR}{1 - FR}$$

and thus

$$\Delta P(p_T, \eta) = \left(\frac{\Delta N_{F,\overline{T}}}{N_{F,\overline{T}}} + \frac{\Delta FR}{FR} + \frac{\Delta FR}{1 - FR} \right) \times N_{F,\overline{T}} \frac{FR}{1 - FR}$$

Error computation for fake rate yields

- Now, for each element in the sum (i.e., for each bin in (p_T, η)), we add the $\Delta P(p_T, \eta)$

$$\Delta N_{\text{fake-}\tau} = \sum_{p_T,\eta} \left[\left(\frac{\Delta N_{F,\overline{T}}}{N_{F,\overline{T}}} + \frac{\Delta FR}{FR} + \frac{\Delta FR}{1-FR} \right) \times N_{F,\overline{T}} \frac{FR}{1-FR} \right]$$