



Study of non-gaussian tails in jet response

Maren Stratmann, Bergische Universität Wuppertal, Germany

September 4, 2019

Abstract

The non-gaussian tails of the asymmetry distribution in di-jet systems are determined for MC simulation and data and compared to each other. The tails are measured separately for b-tagged and non b-tagged events.

First supervisor: Christian Sander

Second supervisor: Othmane Rifki

Contents

1. Introduction	1
2. Theory	1
2.1. Multijet background events	1
2.2. Jet response	2
2.3. Di-jet Asymmetry	3
3. Determining the tail ratio	4
3.1. Fitting the histogram	4
3.2. Calculating the tail fraction	4
4. Results	5
4.1. Statistical limitations	5
4.2. Resolution ratio	7
4.3. Tail ratio	8
5. Conclusion	8
A. Appendix	10

1. Introduction

The non-gaussian tails in jet response are a result of missing transverse momentum in multijet events and have to be modelled correctly in order to search for new physics in events with jets and missing transverse energy. This report describes how the non-gaussian tails in MC simulations were compared with the non-gaussian tails in data. Section 2 provides a short overview over the physical background and important observables. In Section 3 the way the tails were calculated is described and the results of the analysis are presented in Section 4.

2. Theory

2.1. Multijet background events

Multijet events make up a large part of background events for certain signal processes, as for example the decay of SUSY-particles. For the process shown in Figure 1a, the experimental signature would consist of four jets produced by the b-quarks and a fraction of missing transversal energy (MET) from the $\tilde{\chi}^0$ particles. The multijet event shown in Figure 1b can yield a similar experimental signature, as an incorrect measurement of one of the jets also results in a missing transversal energy. Therefore, it is important to understand the multijet background and simulate it correctly in order to search for new physics in events with jets and missing transversal energy in the final state.

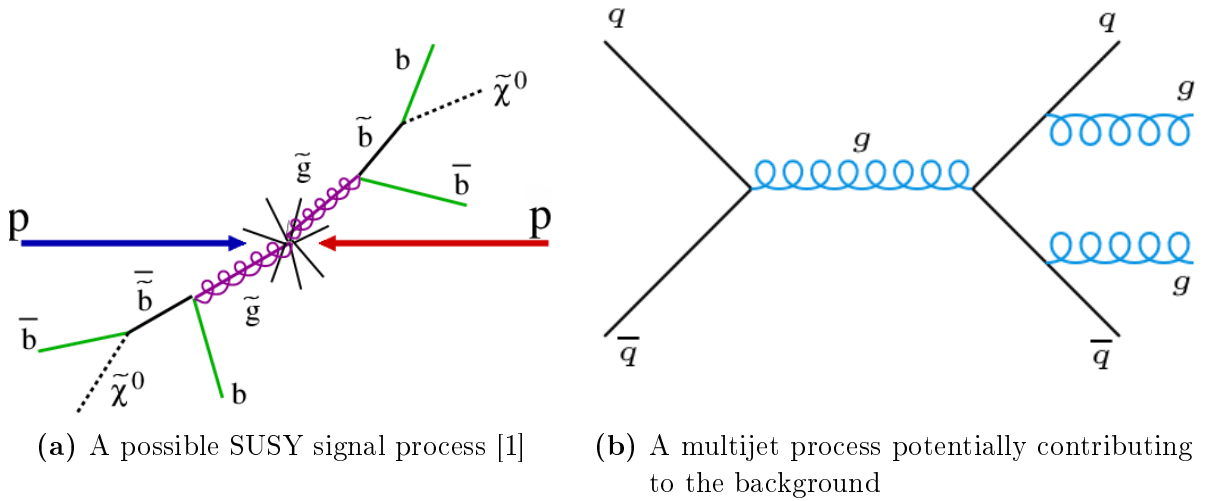


Figure 1: SUSY and multijet events with similar experimental signatures in their final states

2.2. Jet response

The jet response is defined as the reconstructed transversal momentum of a measured jet over its true transversal momentum.

$$p_T = \frac{p_T^{reco}}{p_T^{truth}} \quad (1)$$

As seen in Figure 2, the jet response is gaussian distributed with non-gaussian tails, which are outlined in red and purple. The red part of the tail is a result of missing transverse momentum, which results in a higher fraction of too low p_T^{reco} values than one expects from statistical fluctuations.

It can also be seen that the size of the tail differs for non b-tagged and b-tagged events. Electroweak heavy flavour decays account for a large part of the missing transverse momentum, as neutrinos are produced which can not be measured in the detector. These decays happen more often for b-jets compared to other, lighter jets, which results in a higher fraction of low jet response for b-tagged events.

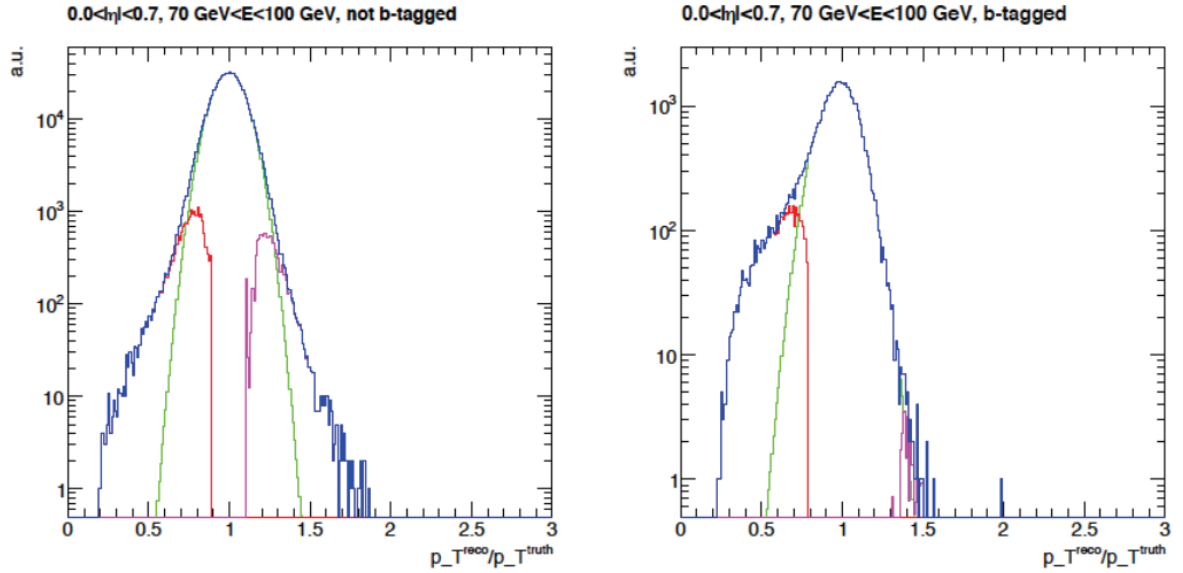


Figure 2: Simulated jet response distribution for b-tagged and non b-tagged events. The green line marks the gaussian distributed fraction and the red and purple lines mark the non-gaussian tails. [2]

2.3. Di-jet Asymmetry

The jet response described in subsection 2.2 can not be measured experimentally, as the actual momentum p_T^{truth} of the jets is not an observable. However, the jet response is connected to the di-jet asymmetry A via the resolution. [2]

$$A = \frac{p_{T,j1} - p_{T,j2}}{p_{T,j1} + p_{T,j2}} \quad (2)$$

$$\frac{\sigma(p_T^{jet})}{\langle p_T^{jet} \rangle} = \sqrt{2}\sigma(A(p_T)) \quad (3)$$

Therefore, it can be assumed that the non-gaussian tails of the asymmetry and the jet response are related as well.

The asymmetry is a measurable quantity which is influenced by additional jet activities, as shown in Figure 3. If there was no additional jet, the asymmetry between jet1 and jet2 would be caused by missing transverse momentum only. With the third jet added to the system, the asymmetry is increased by $\alpha_{||} = \frac{p_{j3||}}{p_{T,ave}}$, which is the part of the momentum of jet3 parallel to the di-jet axis. The mean of the asymmetry distribution depends on $\alpha_{||}$.

$$\langle A \rangle = \frac{p_{T,j1} - p_{T,j2}}{p_{T,j1} + p_{T,j2}} = \frac{p_{j3||}}{2 \cdot p_{T,ave}} = \frac{\alpha_{||}}{2} \quad (4)$$

In Equation 4 it is assumed that there is no missing transverse momentum, as $p_{T,j1} - p_{T,j2} = p_{j3||}$ holds true only then. However, $\langle A \rangle = \frac{\alpha_{||}}{2}$ is a useful approximation for events with missing transverse momentum as well (cf. Table 1).

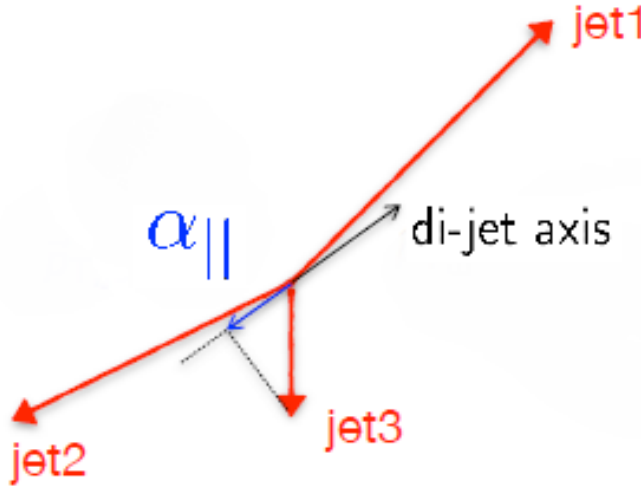


Figure 3: Schematic representation of a jet configuration. Jet1 and jet2 carry the highest transverse momenta. Their di-jet axis is represented by the black arrow. [2]

3. Determining the tail ratio

In the following, the procedure of calculating the non-gaussian tails is described. The starting point for the analysis are 1D histograms generated by ROOT, which show the asymmetry distribution of events (cf. Figure 9). The histograms are produced for different EAve, η and $\alpha_{||}$ bins, where EAve is the average energy of the two jets with highest transverse momentum and η is the pseudorapidity. Histograms with less than 5000 entries were excluded from the analysis due to their low statistics. The fits and calculations described in this section as well as the presented plots were done with ROOT and python.

3.1. Fitting the histogram

A double gaussian is used to fit the histograms to account for non-gaussian behaviour at low asymmetry values (cf. Figure 10).

$$f(x) = h \cdot (e^{(\frac{x-\mu}{\sigma})^2} + e^{(\frac{x+\mu}{\sigma})^2}) \quad (5)$$

h = height of the gaussian

σ = standard deviation

μ = mean value

The mean value μ is fixed at $\frac{\alpha_{||}}{2}$ according to Equation 4. The value of $\alpha_{||}$ is obtained from 2D histograms showing the distribution of events in dependence of $\alpha_{||}$ and the asymmetry by taking the mean value of $\alpha_{||}$ for the respective bin interval (cf. Figure 11). h and σ are used as fit parameters, where the allowed range of σ is restricted to the interval $[0,1]$ in order to prevent negative results.

The height and standard deviation of the histogram as calculated by ROOT are used as initial parameters. The fitting range is restricted to the interval $[\mu-\sigma, \mu+\sigma]$ in order to exclude the non-gaussian tails. After every fit, the fit probability is calculated, which gives the probability of getting a χ^2 value higher than the current one if the fitted function describes the data correctly. If this value is less than 0.1, both borders of the fit interval are increased by one bin and the fit is repeated. If the fit probability is larger than 0.55, both borders of the fit range are decreased by one bin and the fit is repeated as well. A minimal fit range of 20 bins is demanded. A schematic representation of the fit procedure can be found in Figure 4.

3.2. Calculating the tail fraction

The fit results obtained as described in Section 3.1 are used to calculate the tail fractions of the asymmetry distributions. A plot of a histogram with the fit is presented in Figure 5. The tail fraction is defined as the histogram area with a lower limit of $\mu+3\sigma$ normalized to the total area. The choice of the lower limit of the tail is somewhat arbitrary and discussed further in Section 4.2. In Figure 5 the lower tail limit is indicated by the dashed black line. In order to get the non-gaussian tail fraction, the gaussian tail fraction is

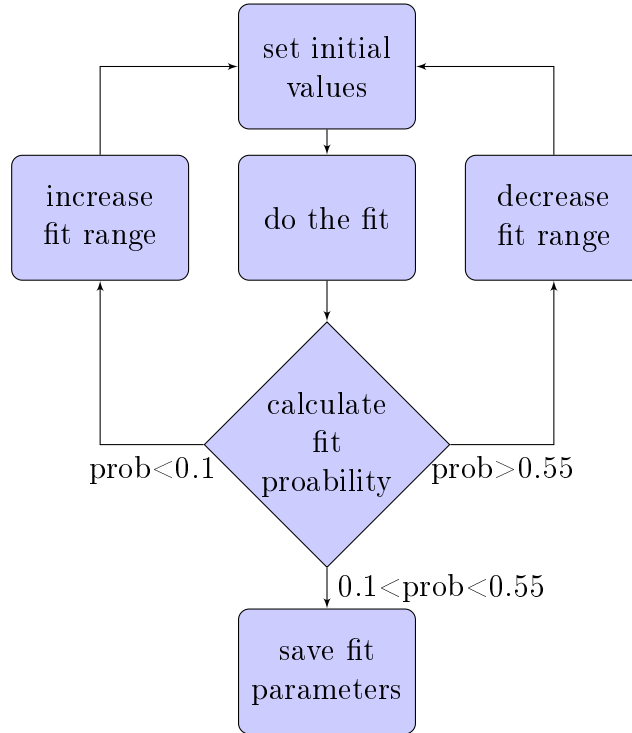


Figure 4: Schematic representation of the fit procedure

calculated and subtracted from the total tail fraction. The gaussian tail fraction is the area beneath the fitted function, again with a lower limit of $\mu+3\sigma$. The fitted function is drawn as the dotted red line in Figure 5.

4. Results

In the following, the tail fractions of the MC simulation and the data are compared to each other. For the MC simulation, the sample mc16a, normalized to 36 fb^{-1} , is used. The data consists of the sets taken with the ATLAS-Detector in 2015 and 2016. The ratios presented in Section 4.2 and Section 4.3 are calculated for an EAve bin of 590-820 GeV and an η bin of 0.0-1.3. The lower limit of the tails is based on the mean value and width of the gaussian fit and is set to $\mu+3\sigma$.

4.1. Statistical limitations

As described in section 3, the tail fraction was only calculated for histograms with at least 5000 entries, as a certain statistic is needed to get meaningful fit results. This condition resulted in some parts of the data being excluded from the calculations, as there are too few events with two b-tagged jets and generally too few events in bins with non-central average energies or high pseudorapidity values as described in the following.

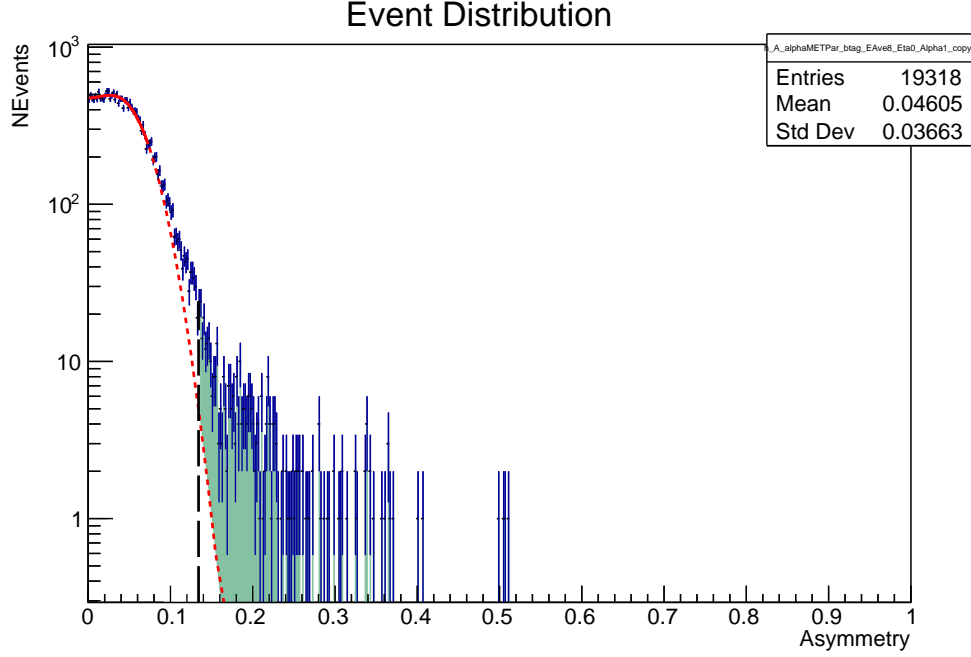


Figure 5: Histogram after the fit procedure. The solid red line is the fit and the dotted red line is the fitted function continued. The dashed black line marks the beginning of the histogram tail at $\mu+3\sigma$. The green area is the non-gaussian tail fraction.

Two b-jets

With the bin sizes initially chosen for the histograms (cf. Table 2), the histograms for events with two b-jets contained too few events to be analysed properly. In order to gain more statistics, the sizes of the E_{Ave} and alpha bins were increased as shown in Table 3. For the simulated data set, these changes did not yield a single histogram for two b-tagged jets with 5000 entries or more. Therefore, the events with two b-tagged jets were excluded from the analysis.

Non-central bins

The amount of events per histogram greatly depends on the E_{Ave} , η and $\alpha_{||}$ bins. Most events are found in central E_{Ave} bins with average energies of roughly 250-820 GeV and low η values from 0.0 to 1.3. The number of events also decreases for higher $\alpha_{||}$ bins. Therefore, most of the asymmetry distribution histograms for the higher η bins and non-central energy bins contained less than 5000 entries and were excluded from the analysis. This effect can be seen more strongly for b-tagged events, as they are generally rarer than non b-tagged events. An example for this can be seen in Figure 12.

4.2. Resolution ratio

As described in subsection 3.2, the tail fractions are calculated with a lower limit of $\mu+3\sigma$, where σ is the fitted width of the gaussian part of the asymmetry distribution. As the tail fraction itself does not depend on the resolution of the gaussian part of the distribution, it has to be established that a resolution dependent lower limit is a reasonable choice. Figure 6 shows the ratio of the resolution values obtained for MC simulation and data. The ratio is close to one for all $\alpha_{||}$ values, which means that the lower limit of the tail fractions based on the resolution values will also be similar enough to make a reasonable starting point. A table with the individual resolution ratios can be found in Table 4.

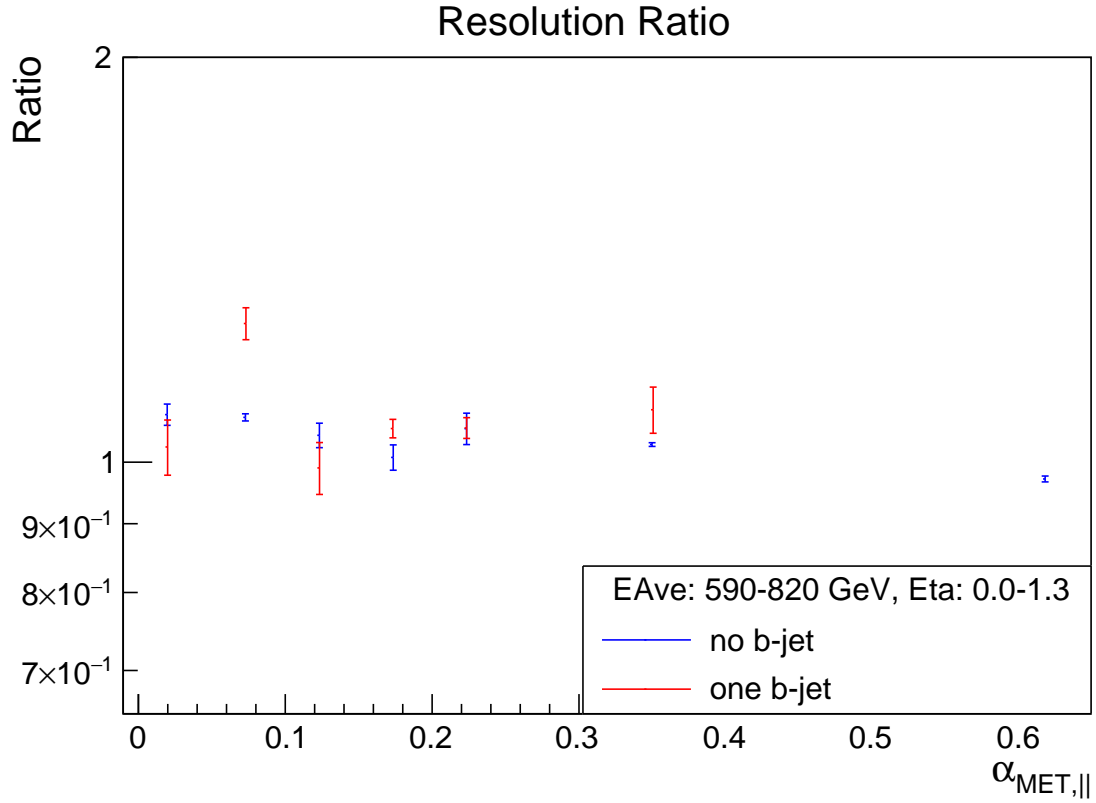


Figure 6: Resolution ratios for b-tagged and non b-tagged events.

4.3. Tail ratio

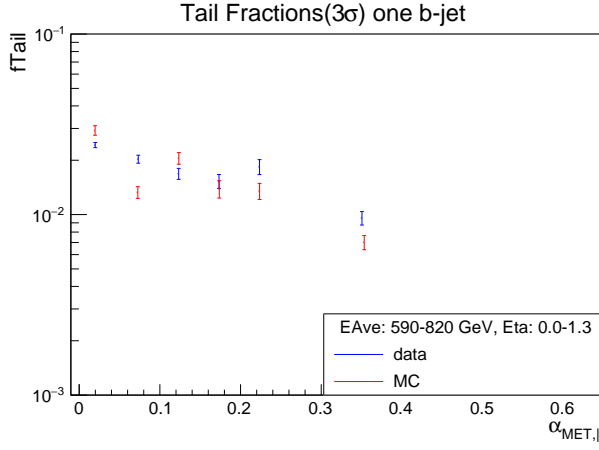
The non-gaussian tail fractions for b-tagged and non b-tagged events are shown in Figure 7. It can be seen that the tail fractions are higher for b-tagged events by roughly a factor of three. As explained in Section 2.2, this was expected because b-jets produce more neutrinos in their final decay states than lighter jets and therefore have a larger probability to be measured with missing transverse momentum.

The tail fractions also show a dependence on $\alpha_{||}$ which is more prominently seen in Figure 7a. For higher $\alpha_{||}$ bins, the uncertainty of the mean value of $\alpha_{||}$ also increases, which leads to wider asymmetry distribution. How the resolution values increase for larger $\alpha_{||}$ values can be seen in Figure 13. A higher resolution shifts the lower end of the tail fraction to higher values, therefore resulting in lower tail fractions. The tail fractions presented in Figure 7 are used to calculate the tail ratio shown in Figure 8. For the tail ratio, the MC tail fraction was divided by the data tail fraction. It can be seen that most of the tail ratios are very close to one, with the b-tagged ratio of 0.65(6) for the second $\alpha_{||}$ value being the largest deviation. The majority of the ratios differs from the expected value of one by less than 20%. A list of all values can be found in Table 5. Overall, it can be said from Figure 8 that the MC simulated non-gaussian tail fractions coincide with the measured ones.

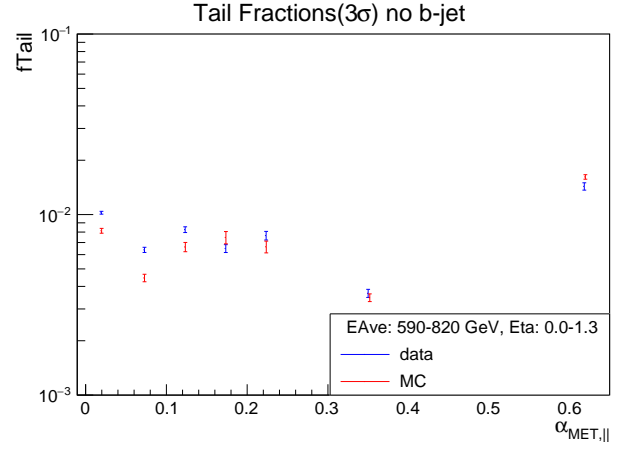
5. Conclusion

The non-gaussian tails in the di-jet asymmetry distribution were calculated and compared for MC simulation and data. There were some statistical limitations to the analysis, as there were not enough events with two b-tagged jets to carry out the analysis. Also, for high η and non-central mean energy values, the statistic of the asymmetry distributions was too low for a proper analysis.

For the mean energy range of 590-820GeV and low η values of 0.0-1.3, the non-gaussian tails in MC simulated data coincide with the non-gaussian tails in measured data. It was also shown that the tail fractions are larger for b-tagged events, as it is to be expected due to b-jets being more difficult to measure.



(a) Tail fractions for b-tagged events



(b) Tail fractions for non b-tagged events

Figure 7: Tail fractions of the gauss fits for MC simulation and data

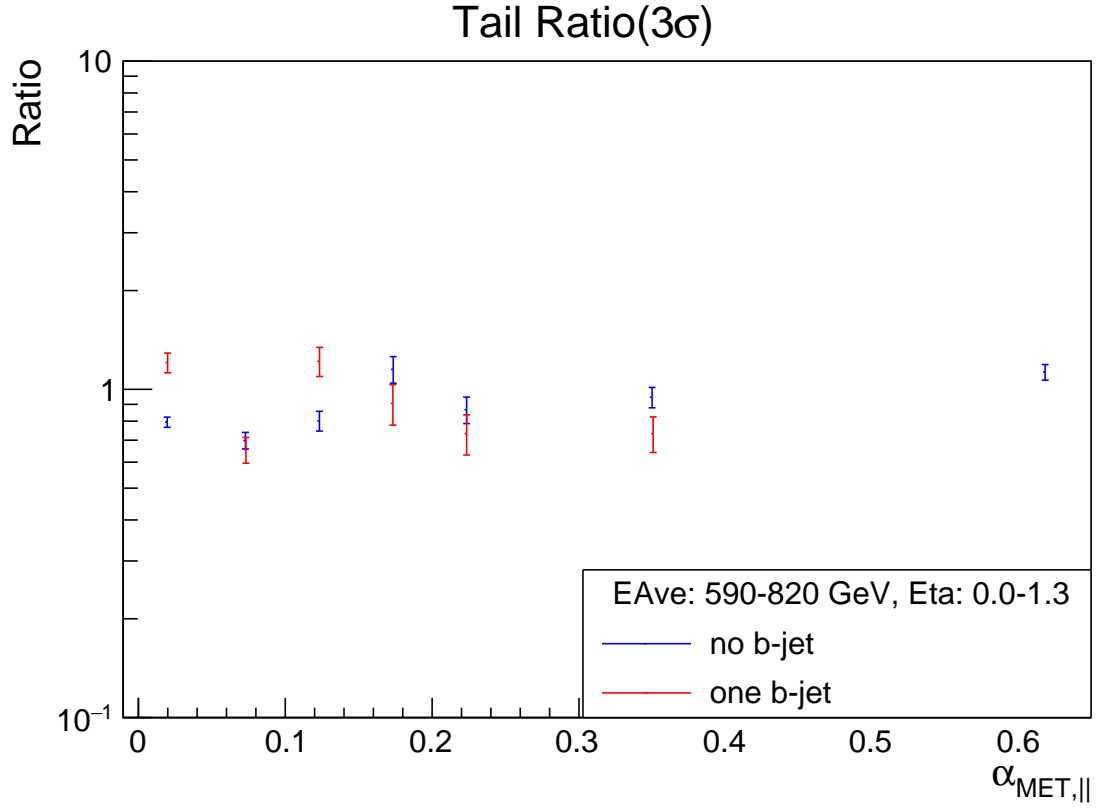


Figure 8: Non-gaussian tail ratios for b-tagged and non b-tagged events.

A. Appendix

Determining the tail ratio

Table 1: Values for $0.5 \cdot \alpha_{||,data}$ and $\langle A \rangle_{data}$ of the distributions for b-tagged and non b-tagged events in different alpha bins for $E_{Ave} = 590-820\text{GeV}$ and $\eta = 0.0-1.3$. The values were acquired with ROOT by getting the mean of the respective histograms. The borders of the alpha bins can be found in Table 2.

	Alpha bin	0	1	2	3	4	5	6
no b-tag	$0.5 \cdot \alpha_{ ,data}$	0.010(7)	0.036(7)	0.0650(2)	0.087(7)	0.175(35)	0.175(35)	0.31(5)
	$\langle A \rangle_{data}$	0.0288(2)	0.0429(8)	0.0645(1)	0.0889(1)	0.1136(2)	0.1764(1)	0.3079(3)
	$0.5 \cdot \alpha_{ ,MC}$	0.010(7)	0.036(7)	0.062(7)	0.07(7)	0.112(7)	0.176(35)	0.31(5)
	$\langle A \rangle_{MC}$	0.0305(8)	0.0441(1)	0.0650(2)	0.0897(2)	0.1145(3)	0.1786(1)	0.3096(3)
b-tag	$0.5 \cdot \alpha_{ ,data}$	0.010(7)	0.037(7)	0.062(7)	0.087(7)	0.112(7)	0.175(35)	-
	$\langle A \rangle_{data}$	0.0325(2)	0.0461(2)	0.0674(4)	0.0901(5)	0.1158(6)	0.1781(4)	-
	$0.5 \cdot \alpha_{ ,MC}$	0.10(7)	0.036(7)	0.062(7)	0.087(7)	0.112(7)	0.177(35)	-
	$\langle A \rangle_{MC}$	0.0347(7)	0.0482(5)	0.0680(6)	0.0913(8)	0.112(7)	0.1792(6)	-

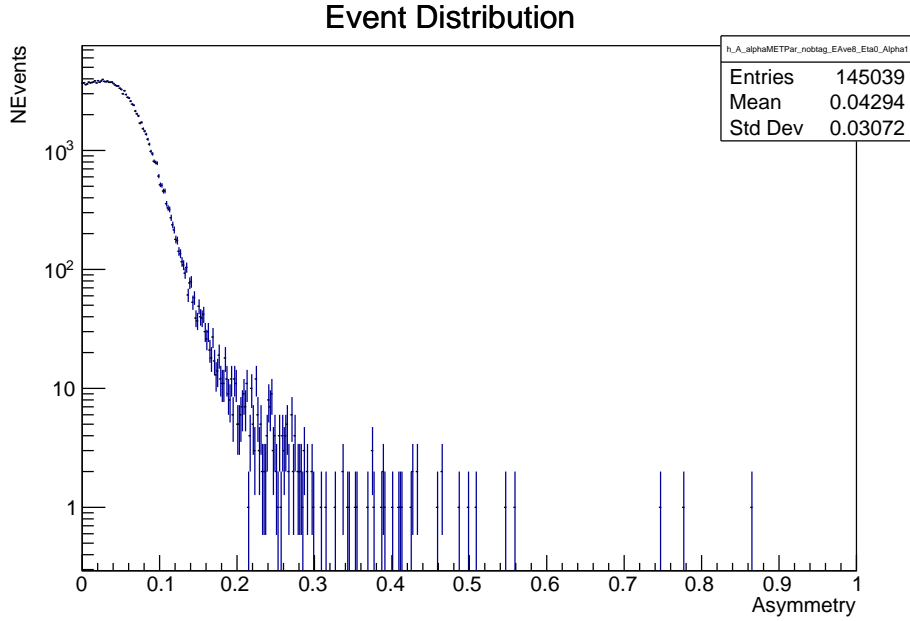


Figure 9: Histogram of the asymmetry distribution for $E_{Ave} = 590-820\text{GeV}$, $\eta = 0.0-1.3$ and $\alpha_{||} = 0.1-0.2$

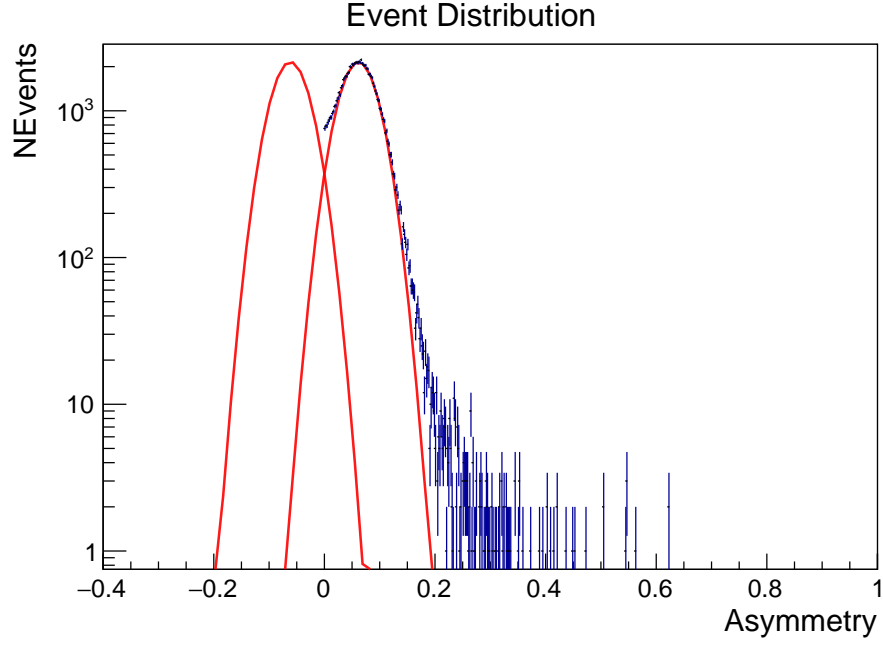


Figure 10: Histogram of the asymmetry distribution for $E_{\text{Ave}} = 590\text{-}820\text{GeV}$, $\eta = 0.0\text{-}1.3$ and $\alpha_{\parallel} = 0.2\text{-}0.25$ with the two gaussians that were used to fit the histogram.

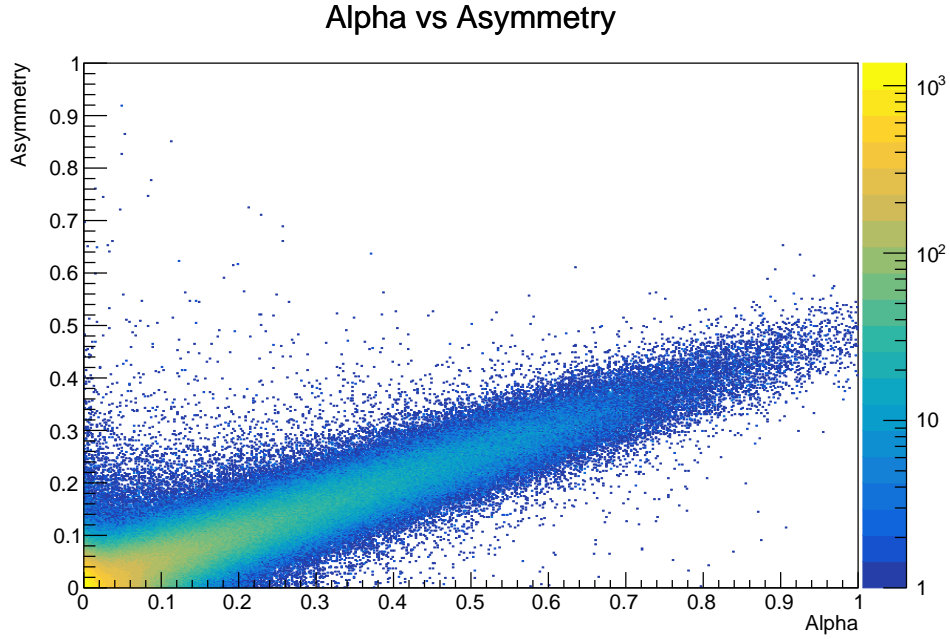


Figure 11: Asymmetry and α_{\parallel} distribution of events for $E_{\text{Ave}} = 590\text{-}820\text{GeV}$ and $\eta = 0.0\text{-}1.3$

Table 2: Boarders of the E_{Ave} , η and alpha bins as used for the histograms analysed in Section 4.

Bin number	0	1	2	3	4	5	6
η	0.0-1.3	1.3-2.5	2.5-5.0				
alpha	0.0-0.05	0.05-0.1	0.1-0.15	0.15-0.2	0.2-0.25	0.25-0.5	0.5-1.0
E_{Ave}	0-20	20-50	50-100	100-190	190-250	250-400	400-590
Bin number	7	8	9	10	11	12	
E_{Ave}	590-820	820-1090	1090-1400	1400-1750	1750-2140	2140-3000	

Table 3: Boarders of the E_{Ave} , η and alpha bins. The E_{Ave} and alpha bins were enlarged to gain more statistics on two b-jet events.

Bin number	0	1	2	3	4	5	6
η	0.0-1.3	1.3-2.5	2.5-5.0				
alpha	0.0-0.1	0.1-0.2	0.2-0.5	0.5-1.0			
E_{Ave}	0-50	50-190	190-400	400-820	820-1400	1400-2140	2140-3000

Results

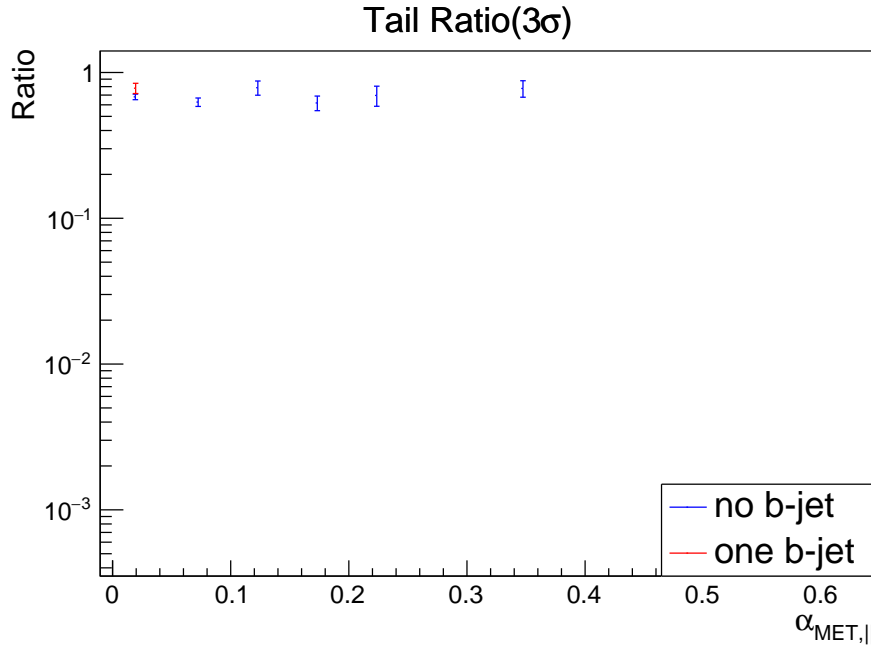


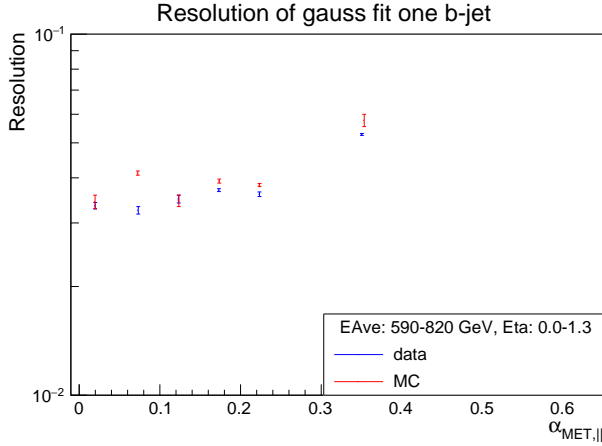
Figure 12: Non-gaussian tail ratios for b-tagged and non b-tagged events for $E_{Ave} = 820-1400\text{GeV}$, $\eta = 0.0-1.3$. The ratio for b-tagged events was only calculated for the first $\alpha_{||}$ bin as the statistics of the other bins were too low.

Table 4: Resolution ratios of MC simulation and data as well as the resolution values for $E_{Ave} = 820\text{-}1400\text{GeV}$ and $\eta = 0.0\text{-}1.3$. The plotted ratios can be found in Figure 8 and the plotted values in Figure 13.

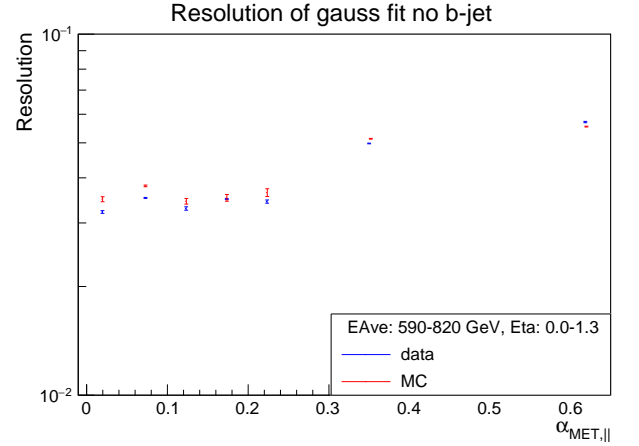
	Point Nr	1	2	3	4	5	6	7
btag	$\alpha_{ }$	0.020(14)	0.073(14)	0.123(14)	0.173(14)	0.224(14)	0.351(70)	-
	Resolution ratio	1.03(5)	1.27(3)	0.99(4)	1.059(17)	1.060(19)	1.094(43)	-
no btag	$\alpha_{ }$	0.0196(14)	0.073(14)	0.123(14)	0.174(14)	0.223(14)	0.350(70)	0.617(99)
	Resolution ratio	1.08(2)	1.080(6)	1.047(22)	1.008(21)	1.059(28)	1.031(3)	0.972(5)

Table 5: Tail ratios for MC simulation and data as well as the tail fractions for $E_{Ave} = 820\text{-}1400\text{GeV}$ and $\eta = 0.0\text{-}1.3$. The plotted ratios can be found in Figure 6 and the plotted fractions in Figure 7.

	Point Nr	1	2	3	4	5	6	7
no b-tag	$\alpha_{ ,data}$	0.020(14)	0.073(14)	0.123(14)	0.173(14)	0.224(14)	0.351(70)	0.617(99)
	tail fraction data	0.0102(2)	0.0064(2)	0.0083(3)	0.0065(3)	0.0076(4)	0.0037(2)	0.0143(7)
	$\alpha_{ ,MC}$	0.02(1)	0.073(1)	0.12(1)	0.17(1)	0.22(1)	0.25(7)	0.62(10)
	tail fraction MC	0.0081(3)	0.0045(2)	0.0066(4)	0.0075(6)	0.0066(5)	0.0035(2)	0.0162(5)
	tail ratio	0.79(3)	0.70(4)	0.80(6)	1.15(11)	0.87(8)	0.95(7)	1.13(6)
b-tag	$\alpha_{ ,data}$	0.0196(14)	0.073(14)	0.123(14)	0.174(14)	0.223(14)	0.350(70)	-
	tail fraction data	0.0242(8)	0.020(1)	0.017(1)	0.015(1)	0.018(2)	0.0096(8)	-
	$\alpha_{ ,MC}$	0.02(1)	0.07(1)	0.12(1)	0.17(1)	0.22(1)	0.35(7)	-
	tail fraction MC	0.029(2)	0.0133(10)	0.021(2)	0.014(2)	0.013(1)	0.0070(6)	-
	tail ratio	1.21(8)	0.65(6)	1.22(12)	0.91(13)	0.73(10)	0.73(9)	-



(a) Resolution values for b-tagged events



(b) Resolution values for non b-tagged events

Figure 13: Resolution values of the gauss fits for MC simulation and data

References

- [1] https://www-cdf.fnal.gov/physics/exotic/r2a/20080410.bbmet_gluinosbottom/. – visited on 28.08.2019
- [2] VINCENT KITALI, CHRISTIAN SANDER : *R+S Uncertainties: Non-Gaussian Tails*