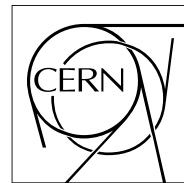


The Compact Muon Solenoid Experiment  
**Analysis Note**

The content of this note is intended for CMS internal use and distribution only



01 April 2013 (v9, 09 April 2014)

# Search for Pair-production of First Generation Scalar Leptoquarks in pp Collisions at $\sqrt{s} = 8$ TeV

E. Berry

*Brown University, Providence, RI, USA*

S. Cooper

*University of Alabama, Tuscaloosa, AL, USA*

P. Rumerio

*University of Alabama, Tuscaloosa, AL, USA*

F. Santanastasio

*CERN, European Organization for Nuclear Research, Geneva, Switzerland*

## Abstract

A search for pair-production of first generation scalar leptoquarks is performed in the final states containing two electrons and at least two jets or an electron, a neutrino, and at least two jets using proton-proton collision data at  $\sqrt{s} = 8$  TeV. The data were collected by the CMS detector at the LHC, corresponding to an integrated luminosity of  $19.6 \text{ fb}^{-1}$ .

# DRAFT

## CMS Physics Analysis Summary

*The content of this note is intended for CMS internal use and distribution only*

2014/04/09

Head Id: 233488

Archive Id: 183956:234854M

Archive Date: 2014/03/25

Archive Tag: trunk

## Search for Pair-production of First Generation Scalar Leptoquarks in pp Collisions at $\sqrt{s} = 8$ TeV

The CMS Collaboration

### Abstract

A search for pair-production of first generation scalar leptoquarks is performed in the final states containing two electrons and at least two jets or an electron, a neutrino, and at least two jets using proton-proton collision data at  $\sqrt{s} = 8$  TeV. The data were collected by the CMS detector at the LHC, corresponding to an integrated luminosity of  $19.6 \text{ fb}^{-1}$ .

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:	Edmund Berry, Paolo Rumerio, Francesco Santanastasio
PDFTitle:	Search for Pair-production of First Generation Scalar Leptoquarks in pp Collisions at $\sqrt{s} = 8$ TeV
PDFSubject:	CMS
PDFKeywords:	CMS, physics, exotica, leptoquark

Please also verify that the abstract does not use any user defined symbols



## 1 Contents

2	1	Introduction . . . . .	3
3	2	Data and Monte Carlo Samples . . . . .	3
4	2.1	Data Samples . . . . .	3
5	2.2	Monte Carlo Samples . . . . .	5
6	3	Reconstruction of Electrons, Muons, Jets, and $E_T^{\text{miss}}$ . . . . .	7
7	3.1	Electrons . . . . .	7
8	3.2	Muons . . . . .	8
9	3.3	Jets and $E_T^{\text{miss}}$ . . . . .	9
10	4	Event Selection . . . . .	10
11	4.1	Event filters . . . . .	10
12	4.2	Trigger . . . . .	10
13	4.3	Event Selection in the $eejj$ Channel . . . . .	12
14	4.4	Event Selection in the $evjj$ Channel . . . . .	16
15	5	Backgrounds . . . . .	22
16	5.1	Other Backgrounds . . . . .	24
17	5.2	QCD Multijet Background . . . . .	25
18	5.3	$t\bar{t}$ Background . . . . .	30
19	5.4	V+jets Background . . . . .	34
20	6	Systematic uncertainties . . . . .	36
21	6.1	Background normalization . . . . .	36
22	6.2	$t\bar{t}$ , $Z^0$ +jets, W+jets background shape . . . . .	36
23	6.3	Electron, jet, and $E_T^{\text{miss}}$ energy scales . . . . .	37
24	6.4	Electron and jet energy resolution . . . . .	37
25	6.5	Integrated luminosity . . . . .	38
26	6.6	MC statistics . . . . .	38
27	6.7	Electron trigger, reconstruction, identification, and isolation uncertainties . . . . .	38
28	6.8	Parton distribution function (PDF) . . . . .	39
29	6.9	Pileup . . . . .	39
30	7	Results . . . . .	44
31	A	$eejj$ analysis $M(LQ) = 650$ final selection extra plots . . . . .	49
32	B	$eejj$ analysis $M(LQ) = 650$ final selection electron quality plots . . . . .	51
33	C	$eejj$ analysis $M(LQ) = 650$ final selection invariant mass plots . . . . .	56
34	D	$eejj$ analysis $M(LQ) = 650$ run period dependence . . . . .	57
35	E	$evjj$ analysis $M(LQ) = 650$ final selection extra plots . . . . .	57
36	F	$evjj$ analysis $M(LQ) = 650$ final selection electron quality plots . . . . .	60
37	G	$evjj$ analysis $M(LQ) = 650$ final selection invariant mass plots . . . . .	62
38	H	$evjj$ analysis $M(LQ) = 650$ run period dependence . . . . .	64
39	I	Results without systematics . . . . .	64
40	J	Comparison with LQ2 . . . . .	64
41	K	Reweighting $E_T^{\text{miss}}$ and $m_{T, ev}$ in the $evjj$ analysis . . . . .	67
42	K.1	Procedure . . . . .	67
43	K.2	Results at preselection . . . . .	72

44	K.3	Results at final selection . . . . .	72
45	L	Purely data-driven background estimate using muons . . . . .	72
46	L.1	Procedure for the $eejj$ analysis . . . . .	75
47	L.2	Results at $eejj$ preselection . . . . .	76
48	L.3	Results at $eejj$ final selection . . . . .	77
49	L.4	Procedure for the $evjj$ analysis . . . . .	80
50	L.5	Results at $evjj$ preselection . . . . .	81
51	L.6	Results at $evjj$ final selection . . . . .	86
52	L.7	Results and conclusions . . . . .	86

DRAFT

## 53 1 Introduction

54 The structure of the standard model (SM) of particle physics suggests a fundamental relationship  
 55 between quarks and leptons. In some theories beyond the SM, such as SU(5) grand-unification [1], Pati-Salam SU(4) [2], composite models [3], technicolor [4–6], and superstring-inspired  $E_6$  models [7], the existence of a new symmetry relates the quarks and leptons in a fundamental way.

59 These models predict the existence of new bosons, called leptoquarks. The leptoquark (LQ) is  
 60 colored, has fractional electric charge, can be either a scalar or vector particle, and couples to a  
 61 lepton and a quark with coupling strength  $\lambda$ . The leptoquark decays to a charged lepton and a  
 62 quark, with unknown branching fraction,  $\beta$ , or a neutrino and a quark, with branching fraction  
 63  $(1 - \beta)$ . Prior to this search, first generation leptoquarks with mass less than 830 (640) GeV had  
 64 been excluded for  $\beta = 1(0.5)$  at 95% confidence level [8].

65 A review of LQ phenomenology and searches can be found in Refs. [9, 10]. Constraints from  
 66 experiments sensitive to flavor-changing neutral currents, lepton-family-number violation, and  
 67 other rare processes favor LQs that couple to quarks and leptons within the same SM generation,  
 68 for LQ masses accessible to current colliders [11]. The dominant mechanisms for the  
 69 production of LQ pairs at the Large Hadron Collider (LHC) are gluon-gluon ( $gg$ ) fusion and  
 70 quark-antiquark ( $q\bar{q}$ ) annihilation, shown in Figure 1.

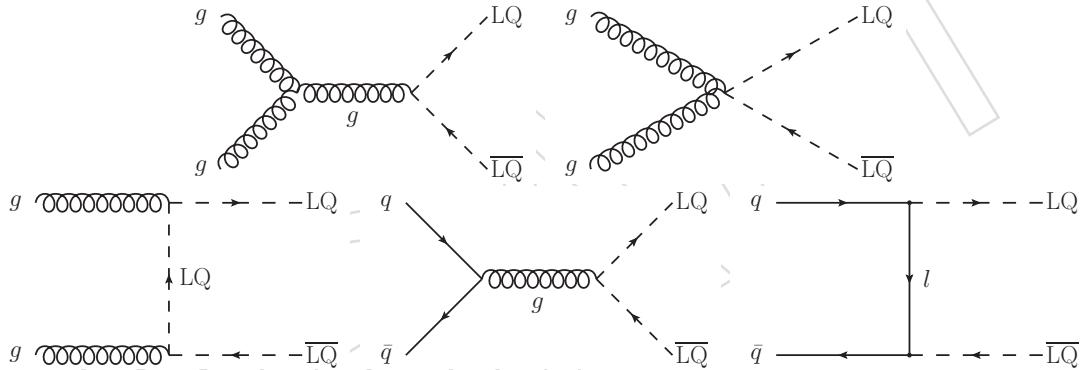


Figure 1: Dominant leading order diagrams for the pair production of scalar leptoquarks.

71 This paper presents the results of a search for pair-production of first generation scalar LQs  
 72 using events containing two electrons and at least two jets ( $eejj$ ) or one electron, missing transverse energy ( $E_T^{\text{miss}}$ ), and at least two jets ( $evjj$ ) using proton-proton collision data at  $\sqrt{s} =$   
 73 8 TeV. In proton-proton collisions at the LHC, LQs are predominantly pair-produced via gluon-  
 74 gluon fusion with a cross section that depends on the strong coupling constant  $\alpha_s$  but is nearly  
 75 independent on  $\lambda$ . The results of this study are based on the assumption that  $\lambda$  is sufficiently  
 76 small that single-LQ production can be neglected. The data were collected in 2012 by the CMS  
 77 detector at the CERN LHC and correspond to an integrated luminosity of  $19.6 \text{ fb}^{-1}$ .

## 79 2 Data and Monte Carlo Samples

### 80 2.1 Data Samples

81 The datasets used for this analysis were collected by the CMS detector during 2012 with the  
 82 LHC operating at  $\sqrt{s} = 8$  TeV and correspond to  $19.6 \text{ fb}^{-1}$ . Only data that were taken during  
 83 a period with stable LHC beams and with all CMS sub-detectors operating without problems

- have been used in the analysis<sup>1</sup>. Known dead or noisy channels in the calorimeters were masked in the official reconstruction.
- For the LQ searches in both the  $eejj$  channel and the  $evjj$  channel, the data stored in the *ElectronHad* Primary Dataset (PD) are used as reported in Table 1. The events used in this channel are collected by unprescaled electron + dijet triggers as described in Section 4.2. For the de-

Table 1: *ElectronHad* datasets, the run range considered, and the corresponding integrated luminosity,  $\mathcal{L}_{int}$ , of the sample after removing data taking periods with known detector problems.

Dataset	Run range	$\mathcal{L}_{int}(\text{pb}^{-1})$
/ElectronHad/Run2012A-recover-06Aug2012-v1/AOD	190782 - 190949	82
/ElectronHad/Run2012A-13Jul2012-v1/AOD	190645 - 193621	808
/ElectronHad/Run2012B-13Jul2012-v1/AOD	193834 - 196531	4430
/ElectronHad/Run2012C-24Aug2012-v1/AOD	198049 - 198522	495
/ElectronHad/Run2012C-EcalRecover_11Dec2012-v1/AOD	201191	134
/ElectronHad/Run2012C-PromptReco-v2/AOD	198941 - 203002	6390
/ElectronHad/Run2012D-PromptReco-v1/AOD	203894 - 208686	7270
Total integrated luminosity ( $\mathcal{L}_{int}$ )	19.6 $\text{fb}^{-1}$	

- termination of the QCD multijet background in both the  $eejj$  channel and the  $evjj$  channel, the data stored in the *Photon* and *SinglePhoton* PDs are used as reported in Table 2. The events used in this study are collected by prescaled single-photon triggers as described in Section 5.2. For

Table 2: *Photon* and *SinglePhoton* datasets, the run range considered, and the corresponding integrated luminosity,  $\mathcal{L}_{int}$ , of the sample after removing data taking periods with known detector problems.

Dataset	Run range	$\mathcal{L}_{int}(\text{pb}^{-1})$
/Photon/Run2012A-recover-06Aug2012-v1/AOD	190782 - 190949	82
/Photon/Run2012A-13Jul2012-v1/AOD	190645 - 193621	808
/SinglePhoton/Run2012B-13Jul2012-v1/AOD	193834 - 196531	4430
/SinglePhoton/Run2012C-24Aug2012-v1/AOD	198049 - 198522	495
/SinglePhoton/Run2012C-EcalRecover_11Dec2012-v1/AOD	201191	134
/SinglePhoton/Run2012C-PromptReco-v2/AOD	198941 - 203002	6390
/SinglePhoton/Run2012D-PromptReco-v1/AOD	203894 - 208686	7270
Total integrated luminosity ( $\mathcal{L}_{int}$ )	19.6 $\text{fb}^{-1}$	

- the determination of the  $t\bar{t}$  background in the  $eejj$  channel only, the data stored in the *SingleMu* PDs are used as reported in Table 3. The events used in this study are collected by unprescaled single-muon triggers as described in Section 5.3.1.

<sup>1</sup>JSON files:

/afs/cern.ch/cms/CAF/CMSCOMM/COMM.DQM/certification/Collisions12/8TeV/Reprocessing/Cert.190456-196531.8TeV\_13Jul2012ReReco.Collisions12.JSON.v2.txt  
/afs/cern.ch/cms/CAF/CMSCOMM/COMM.DQM/certification/Collisions12/8TeV/Reprocessing/Cert.190782-190949.8TeV\_06Aug2012ReReco.Collisions12.JSON.txt  
/afs/cern.ch/cms/CAF/CMSCOMM/COMM.DQM/certification/Collisions12/8TeV/Reprocessing/Cert.198022-198523.8TeV\_24Aug2012ReReco.Collisions12.JSON.txt  
/afs/cern.ch/cms/CAF/CMSCOMM/COMM.DQM/certification/Collisions12/8TeV/Reprocessing/Cert.201191-201191.8TeV\_11Dec2012ReRecorecover.Collisions12.JSON.txt  
/afs/cern.ch/cms/CAF/CMSCOMM/COMM.DQM/certification/Collisions12/8TeV/Prompt/Cert.190456-208686.8TeV\_PromptReco.Collisions12.JSON.txt

Table 3: *SingleMu* datasets, the run range considered, and the corresponding integrated luminosity,  $\mathcal{L}_{int}$ , of the sample after removing data taking periods with known detector problems.

Dataset	Run range	$\mathcal{L}_{int}(\text{pb}^{-1})$
/SingleMu/Run2012A-recover-06Aug2012-v1/AOD	190782 - 190949	82
/SingleMu/Run2012A-13Jul2012-v1/AOD	190645 - 193621	808
/SingleMu/Run2012B-13Jul2012-v1/AOD	193834 - 196531	4430
/SingleMu/Run2012C-24Aug2012-v1/AOD	198049 - 198522	495
/SingleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD	201191	134
/SingleMu/Run2012C-PromptReco-v2/AOD	198941 - 203002	6390
/SingleMu/Run2012D-PromptReco-v1/AOD	203894 - 208686	7270
Total integrated luminosity ( $\mathcal{L}_{int}$ )	$19.6 \text{ fb}^{-1}$	

## 2.2 Monte Carlo Samples

The dominant sources of  $eejj$  and  $evjj$  events from production of standard model particles are pair-production of top quarks and associated production of a  $Z^0$  or  $W$  boson with jets. Smaller background contributions come from QCD multijet events, single top, diboson, and  $\gamma+jets$  production.

To compare data to Monte Carlo (MC), the response of the detector was simulated using GEANT [12, 13]. The detector geometry description included realistic subsystem conditions such as dead and noisy channels. The selection procedure as well as the electron, muon, jet, and  $E_T^{\text{miss}}$  reconstructions described for the data in Section 3 are also applied to the MC simulation samples. The trigger selection is only applied to the data. The effect of this approach on the results is studied for both  $eejj$  and  $evjj$  analyses in Section 6.7. The presence of pileup events is included in the simulation (a re-weighting of MC events for pileup is applied to match with the number of pileup events observed in data, accordingly with the CMS prescription). For the generation of all the MC samples the CTEQ6L [14] parton distribution functions (PDFs) were used. The  $W+jets$ ,  $Z^0+jets$ , and  $t\bar{t}$  events were generated using MADGRAPH [15]. The  $\gamma+jets$  and diboson ( $WW$ ,  $WZ$ ,  $ZZ$ ) events were generated using PYTHIA [16], version 6.422. The single-top events were generated using POWHEG [17]. For the MADGRAPH and POWHEG samples, parton showering and hadronization were performed with PYTHIA. The QCD multijet background is estimated from data for both the  $eejj$  and  $evjj$  analyses, as described in Section 5.2. The  $t\bar{t}$  background is estimated from data for the  $eejj$  analysis only, as described in Section 5.3.1.

The total MADGRAPH cross section for the  $W+jets$  ( $Z^0+jets$ ) is rescaled to an inclusive NNLO  $W \rightarrow \ell\nu$  ( $Z/\gamma \rightarrow \ell\ell$ ) production cross section of 37509.0 pb (3503.71 pb), calculated using FEWZ [18], where  $\ell = e, \mu$ , or  $\tau$ . The MADGRAPH  $t\bar{t}$  sample is normalized to an inclusive NNLO cross section of 234 pb calculated in Ref [19]. For single-top samples, NNLO cross sections of 5.55 pb for the s-channel, 87.1 pb for the t-channel, and 22.2 pb for the tW-channel, respectively, are used [19]. The  $WW$ ,  $WZ$ , and  $ZZ$  samples are normalized to the NLO cross sections of 57.1 pb, 32.3 pb, and 8.26 pb, respectively, calculated with NNPDF. For the  $\gamma+jets$  samples, LO cross sections calculated with PYTHIA are used. The MC dataset path and cross section for all background processes encountered in this analysis are listed in Table 4.

Signal samples for leptoquark masses from 300 to 1200 GeV were generated with PYTHIA for both  $eejj$  ( $LQL\bar{Q} \rightarrow eqe\bar{q}$ ) and  $evjj$  ( $LQL\bar{Q} \rightarrow eq\nu\bar{q}'$ ) final states. All samples were generated

Table 4: Monte Carlo datasets used to model SM backgrounds, the simulated process, and the corresponding cross section.

Dataset	Process	Cross section [pb]
/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/	W+ jets	37509.0
/W1JetsToLNu_TuneZ2Star_8TeV-madgraph/		6601.92
/W2JetsToLNu_TuneZ2Star_8TeV-madgraph/		2107.48
/W3JetsToLNu_TuneZ2Star_8TeV-madgraph/		634.10
/W4JetsToLNu_TuneZ2Star_8TeV-madgraph/		260.41
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/	Z/ $\gamma^*$ + jets	3503.71
/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/		666.30
/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/		214.97
/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/		60.69
/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/		27.36
/G_Pt-15to30_TuneZ2star_8TeV_pythia6/	$\gamma$ + jets	200062
/G_Pt-30to50_TuneZ2star_8TeV_pythia6/		19932
/G_Pt-50to80_TuneZ2star_8TeV_pythia6/		3322.3
/G_Pt-80to120_TuneZ2star_8TeV_pythia6/		558.3
/G_Pt-120to170_TuneZ2star_8TeV_pythia6/		108.0
/G_Pt-170to300_TuneZ2star_8TeV_pythia6/		30.12
/G_Pt-300to470_TuneZ2star_8TeV_pythia6/		2.139
/G_Pt-470to800_TuneZ2star_8TeV_pythia6/		0.2119
/G_Pt-800to1400_TuneZ2star_8TeV_pythia6/		0.007078
/G_Pt-1400to1800_TuneZ2star_8TeV_pythia6/		4.510E-5
/G_Pt-1800_TuneZ2star_8TeV_pythia6/		1.867E-6
/TTJets_FullLeptMGDecays_8TeV-madgraph/		26.18
/TTJets_SemiLeptMGDecays_8TeV-madgraph/	$t\bar{t}$	103.71
/TTJets_HadronicMGDecays_8TeV-madgraph/		104.10
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/	Single top	11.1
/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/		30.7
/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/		1.76
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/		11.1
/T_t-channel_TuneZ2star_8TeV-powheg-tauola/		56.4
/T_s-channel_TuneZ2star_8TeV-powheg-tauola/		3.79
/WW_TuneZ2star_8TeV_pythia6_tauola/	WW, WZ, ZZ	57.1
/WZ_TuneZ2star_8TeV_pythia6_tauola/		32.3
/ZZ_TuneZ2star_8TeV_pythia6_tauola/		8.26

<sup>127</sup> with  $\lambda = 0.3$ . Table 5 shows the leptoquark pair production NLO cross section and the relative  
<sup>128</sup> theoretical uncertainties as a function of the leptoquark mass [20].

Table 5: Details of the Monte Carlo signal datasets used in the analysis. Table includes NLO cross sections [pb], PDF uncertainty, and NLO cross-sections with the renormalisation/factorization scale varied between half and twice the leptoquark mass [20].

$M_{LQ}$	$\sigma(\mu = M_{LQ})$	$\delta(PDF)$	$\sigma(\mu = M_{LQ}/2)$	$\sigma(\mu = M_{LQ} \times 2)$
300	1.89	0.214	1.63	2.13
350	0.77	0.102	0.663	0.866
400	0.342	0.052	0.295	0.385
450	0.163	0.0278	0.14	0.183
500	0.082	0.0155	0.0704	0.0922
550	0.0431	0.00893	0.037	0.0485
600	0.0235	0.0053	0.0201	0.0265
650	0.0132	0.00322	0.0113	0.0149
700	0.00761	0.002	0.00648	0.00858
750	0.00448	0.00126	0.00381	0.00506
800	0.00269	0.00081	0.00228	0.00304
850	0.00164	0.000527	0.00139	0.00186
900	0.00101	0.000347	0.000856	0.00115
950	0.000634	0.000231	0.000534	0.000722
1000	0.000401	0.000155	0.000337	0.000458
1050	0.000256	0.000105	0.000214	0.000293
1100	0.000165	7.18e-05	0.000138	0.000189
1150	0.000107	4.92e-05	8.88e-05	0.000123
1200	6.96e-05	3.4e-05	5.77e-05	8.04e-05

### <sup>129</sup> 3 Reconstruction of Electrons, Muons, Jets, and $E_T^{\text{miss}}$

#### <sup>130</sup> 3.1 Electrons

<sup>131</sup> Electron candidates (“GsfElectron” collection), reconstructed offline, are required to have an  
<sup>132</sup> electromagnetic (EM) cluster with transverse energy ( $E_T$ )  $> 35$  GeV and pseudo-rapidity  $|\eta| <$   
<sup>133</sup> 2.5, excluding the transition region between the barrel and endcap detectors,  $1.442 < |\eta| <$   
<sup>134</sup> 1.56. The “HEEP v4.1” selection criteria for electron ID and isolation [21], which are optimized  
<sup>135</sup> for electrons with energies of hundreds of GeV and are commonly used in the Exotica group,  
<sup>136</sup> have been applied and are described in Table 6. The definitions of these cuts (except  $d_{xy}$ ) can  
<sup>137</sup> be found in Ref [22].

<sup>138</sup> The electron reconstruction, identification, and isolation efficiencies have been studied in the  
<sup>139</sup> context of the  $Z' \rightarrow ee$  analysis [21]. A data/MC scale factor for the single electron efficiency  
<sup>140</sup> equal to 0.98 (with a relative uncertainty of 1.5%) has been measured for events with dielectron  
<sup>141</sup> invariant mass greater than 140 GeV, i.e. above the  $Z$  peak (similarly to the  $eejj$  events passing  
<sup>142</sup> the full selection of this search). The 0.98 ( $0.98 \times 0.98 = 0.96$ ) scale factor reported by the  
<sup>143</sup>  $Z' \rightarrow ee$  analysis is applied to the  $evjj$  ( $eejj$ ) signal predictions in the form of 2% (4%) systematic  
<sup>144</sup> uncertainties, since those are derived from MC. No scale factors or systematic uncertainties  
<sup>145</sup> related to this effect are applied to the major backgrounds, since their normalization is already  
<sup>146</sup> derived from comparison with data in control regions.

Table 6: The “HEEP v4.1” selection criteria for electron ID and isolation [21]. Note that calorimeter isolation corresponds to EM isolation + Hadronic depth 1 isolation.

Variable	Barrel	Endcap
$E_T$ [ GeV ]	$> 35$	$> 35$
$ \eta_{SC} $	$ \eta  < 1.442$	$1.56 <  \eta  < 2.5$
seed	ECAL seeded	ECAL seeded
Missing hits	$\leq 1$	$\leq 1$
$ d_{xy} $ [ cm ]	$< 0.02$	$< 0.05$
$\Delta\eta_{in}$	$< 0.005$	$< 0.007$
$\Delta\phi_{in}$	$< 0.06$	$< 0.06$
$H/E$	$< 0.05$	$< 0.05$
$\sigma_{i\eta i\eta}$	-	$< 0.03$
Shape	$E^{2\times 5}/E^{5\times 5} > 0.94$ OR $E^{1\times 5}/E^{5\times 5} > 0.83$	-
Calo iso. [ GeV ]	$< 2 + 0.03 \times E_T + 0.28 \times \rho$	$\begin{cases} < 2.5 + 0.28 \times \rho & \text{if } E_T < 50. \\ < 2.5 + 0.28 \times \rho + 0.03 \times (E_T - 50.) & \text{if } E_T \geq 50. \end{cases}$
Tracker iso. [ GeV ]	$< 5.0$	$< 5.0$

## 3.2 Muons

Muons are employed in both the  $eejj$  and the  $evjj$  channels, where a veto on the presence of isolated muons is used to suppress  $t\bar{t}$  backgrounds. In addition, muons are employed in the  $eejj$  channel for the  $t\bar{t}$  background estimation using the  $e\mu jj$  control sample.

Muon candidates are reconstructed as tracks in the muon system that are matched to a track that has been reconstructed by the inner tracking system [23]. For this analysis, muons must have  $p_T > 10$  GeV and satisfy the tight muon ID requirements, which include the following:

- Identified as a global muon;
- Passed the particle flow muon ID;
- $\chi^2/\text{NDOF}$  of the global muon track fit  $< 10$ ;
- At least one muon chamber hit included in global muon track fit;
- At least one pixel hit included in the inner track hit pattern;
- At least six tracker layers have hits;
- Muon segments must be found in at least two muon stations;
- Tracker track has transverse impact parameter  $d_{xy} < 2$  mm with respect to the primary vertex;
- Tracker track has longitudinal impact parameter  $d_z < 5$  mm with respect to the primary vertex;

In addition, muons must be isolated by requiring the combined relative PF-based isolation within a cone of 0.4 from the muon be less than 0.12. This isolation is corrected for pileup using the “DeltaB” corrections.

Combined relative PF-based isolation ( $I_{\text{total}}$ ) is defined by the relation in Equation 1:

$$I_{\text{total}} = \frac{I_{\text{charged hadron}} + \max(0., I_{\text{neutral hadron}} + I_{\text{photon}} - \frac{1}{2} \cdot I_{\text{pileup}})}{p_T} \quad (1)$$

where  $I_{\text{charged hadron}}$  refers to the sum  $p_T$  of the particle-flow identified charged hadrons within the isolation cone,  $I_{\text{neutral hadron}}$  refers to the sum  $p_T$  of the particle-flow identified neutral hadrons within the isolation cone,  $I_{\text{photon}}$  refers to the sum  $p_T$  of the particle-flow identified photons within the isolation cone, and  $I_{\text{pileup}}$  refers to the sum  $p_T$  of the particle-flow identified charged particles within the isolation cone that do not originate at the primary vertex.

### 3.3 Jets and $E_T^{\text{miss}}$

Jets are reconstructed by the anti- $k_T$  algorithm [24] from a list of particles obtained using particle-flow methods and a radius parameter  $R = 0.5$ . The particle-flow algorithm [25, 26] reconstructs a complete, unique list of particles in each event using an optimized combination of information from all CMS subdetector systems. Particles that are reconstructed and identified include muons, electrons (with associated bremsstrahlung photons), photons (unconverted and converted), and charged and neutral hadrons.

The jet energy scale corrections are derived using Monte Carlo (MC) simulation and *in situ* measurements using dijet and photon+jet events. The following corrections are applied: “L1FastJet” (pileup), “L2Relative”, “L3Absolute”, and “L2L3Residual”. The latter residual corrections are applied only to jets in data. The corrections are extracted from a global tag according to the dataset being analyzed. The list of datasets and their corresponding global tags may be found in Table 7.

Table 7: List of the global tags used to analyze all datasets in this analysis

Dataset	GlobalTag
All Monte Carlo datasets	START53_V7G
Run2012A-recover-06Aug2012-v1	FT_53_V6C_AN3
Run2012A-13Jul2012-v1	FT_53_V6C_AN3
Run2012B-13Jul2012-v1	FT_53_V6C_AN3
Run2012C-24Aug2012-v1	FT53_V10A_AN3
Run2012C-EcalRecover_11Dec2012-v1	FT_P_V42C_AN3
Run2012C-PromptReco-v2	GR_P_V42_AN3
Run2012D-PromptReco-v1	GR_P_V42_AN3

In addition to the PF calorimeter noise cleaning at reconstruction level, a set of cuts referred to as *loose jet ID* [27, 28] is applied to further remove fake jets due to electronic noise or other detector artifacts:

- Neutral hadron (EM) fraction  $< 0.99$ , for all jet  $\eta$ ;
- Charged hadron (EM) fraction  $> 0 (< 0.99)$ , for jet  $|\eta| < 2.4$ ;
- Number of constituents  $> 1$ , for all jet  $\eta$ ;
- Charged multiplicity  $> 0$ , for jet  $|\eta| < 2.4$ ;

Finally, to address the double-counting of well identified and isolated electrons and muons reconstructed also as jets, for each selected electron or muon, the closest selected jet within  $\Delta R = 0.3$  is removed from the jet collection.

197 The missing transverse energy,  $E_T^{\text{miss}}$ , of the event is computed as the negative vector sum of all  
 198 particle flow objects' transverse momenta. Corrections are applied to the  $E_T^{\text{miss}}$  to account for  
 199 pileup ("type-0"), jet energy corrections ("type-1"), and a systematic shift of the  $E_T^{\text{miss}}$  observed  
 200 in the  $x - y$  plane. More information about  $E_T^{\text{miss}}$  performance during this running period can  
 201 be found in Ref [29].

## 202 4 Event Selection

### 203 4.1 Event filters

204 All events used in both the  $eejj$  and  $evjj$  analyses are required to pass the following standard  
 205 event filters (applied offline), in order to reject beam background, electronic noise, and other  
 206 detector artifacts:

- 207 • HBHE noise filter : The recommended version for CMSSW\_5\_3\_8 is used;
- 208 • HCAL laser filter : events containing HCAL calibration firing are removed using  
 209 HCAL DIGI information. This information is not available in the AOD-level infor-  
 210 mation that is processed in this analysis, so events have been identified by the HCAL  
 211 POG independently and are filtered from this analysis by means of an event list;
- 212 • ECAL dead cell filters : Fewer than 1% of the crystals in the ECAL are masked, either  
 213 because they have single noisy crystals or because they correspond to front end cards  
 214 with no data link. Energy lost in regions with masked ECAL crystals may lead to  
 215 fake  $E_T^{\text{miss}}$  in the event. These events are identified for removal by using trigger  
 216 primitive information and by using the energy surrounding the masked crystals to  
 217 estimate how much energy was lost;
- 218 • Bad EE supercrystal filter : events with anomalously high energies from two specific  
 219 ECAL endcap 5x5 crystal regions are identified for removal by cutting on the total  
 220 supercluster energy and the number of bad rechits with energy greater than 1 TeV  
 221 in the two superclusters;
- 222 • EB or EE crystals with large laser calibration correction filter : a multiplicative factor  
 223 correcting for crystal transparency loss is applied to ECAL crystal energy measure-  
 224 ments. These constants are usually greater than 1.0. A small number of crystals  
 225 can receive abnormally large (greater than 3.0 in the barrel and 8.0 in the endcap)  
 226 correction values, resulting in fake  $E_T^{\text{miss}}$ .
- 227 • Tracking failure filter : the sumpt of tracks belonging to good vertices divided by the  
 228 HT of all jets in an event is required to be greater than 10%.
- 229 • Beam halo filter : The "tight" version of this filter (CSCTightHaloFilter) is used,  
 230 as recommended for 50ns bunch spacing;
- 231 • Beam scraping filter : the fraction of high-purity tracks is requested to be greater  
 232 than 25% in events with more than 10 tracks. This removes rare beam-induced back-  
 233 ground that produces a high multiplicity of clusters in the pixel detector;
- 234 • Good primary vertex : at least one well-identified primary vertex defined as `!vertex->isFake()`,  
 235 `vertex->ndof() > 4`, `| vertex->z() | <= 24 cm`, and `vertex->position().rho()`  
 236 `<= 2 cm`;

### 237 4.2 Trigger

238 Events are selected online using unprescaled single-electron plus dijet triggers depending on  
 239 the run period, as reported in Tables 8 and 9.

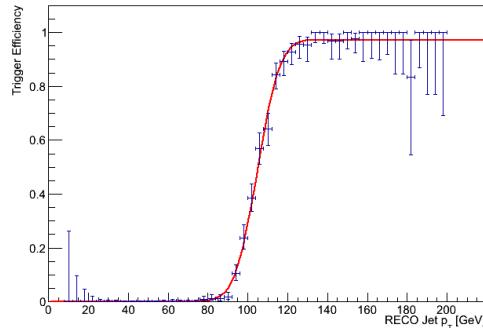
Table 8: Electron + two particle flow jet HLT paths used for signal selection in the analysis.

HLT path	Run range
HLT_Ele30_CaloIdVT_TrkIdT_PFJet100_PFJet25_v3	190456 - 190738
HLT_Ele30_CaloIdVT_TrkIdT_PFJet100_PFJet25_v4	190782 - 191419

Table 9: Electron + two particle flow jet HLT paths used for signal selection in the analysis. The estimated contribution from pileup energy is removed from the jets.

HLT path	Run range
HLT_Ele30_CaloIdVT_TrkIdT_PFNopPUJet100_PFNopPUJet25_v4	191691 - 194225
HLT_Ele30_CaloIdVT_TrkIdT_PFNopPUJet100_PFNopPUJet25_v5	194270 - 196531
HLT_Ele30_CaloIdVT_TrkIdT_PFNopPUJet100_PFNopPUJet25_v6	198022 - 199608
HLT_Ele30_CaloIdVT_TrkIdT_PFNopPUJet100_PFNopPUJet25_v7	199698 - 202504
HLT_Ele30_CaloIdVT_TrkIdT_PFNopPUJet100_PFNopPUJet25_v8	202970 - 208686

- 240 The efficiency for the electron leg of these triggers to select events with a single electron passing  
 241 the HEEP selection criteria has been studied in the context of the  $W' \rightarrow e\nu$  analysis [30], and it  
 242 is taken to be  $97.4 \pm 0.56\%$  ( $95.8 \pm 1.35\%$ ) in the barrel (endcap).  
 243 The jet PFJet100 or PFNopPUJet100 leg of these triggers have been studied in a single-  
 244 electron sample and found to be fully efficient selecting events with an anti- $k_T$  particle flow  
 245 jet passing the loose jet ID and having  $p_T > 125$  GeV (see Figure 2).

Figure 2: Efficiency of the jet-leg of PFJet100 leg of the signal trigger in selecting events with an anti- $k_T$  particle flow jet passing the loose jet ID as a function of offline jet  $p_T$ . An offline cut of  $p_T > 125$  GeV is taken for the analysis preselection. This efficiency was taken from a sample of events passing the prescaled HLT\_Ele30\_CaloIdVT\_TrkIdT trigger and passing the PFJet25 leg of the signal trigger.

- 246 The total trigger efficiency is taken, then, to be equal to the efficiency of the electron leg to select  
 247 events with a single electron passing the HEEP selection criteria. This efficiency is applied to  
 248 the signal Monte Carlo. It is not applied to the major backgrounds since their normalization is  
 249 already derived from comparison with data in control regions.

### 4.3 Event Selection in the $eejj$ Channel

This section describes the event selection in the search for leptoquarks in the decay channel  $LQL\bar{Q} \rightarrow ee\bar{q}\bar{q}$ . The two leading (in  $p_T$ ) electrons and the two leading (in  $p_T$ ) jets are used in this analysis. The definition of some reconstructed quantities that are used in the  $eejj$  analysis are described below:

- $m_{ee}$ , the electron-electron invariant mass;
- $S_T$ , the scalar sum of the  $p_T$  of the two electrons and the two leading jets;
- $m_{ej}$  ( $m_{ej}^{\min}$ ), the average (minimum) electron-jet invariant mass of the two leptoquark candidates, obtained from the two electrons and two jets. There are two possible ways to combine two electrons and two jets to form two leptoquark candidates: (e1-j1,e2-j2) or (e1-j2,e2-j1). The combination with the smaller difference between the two electron-jet masses is considered in this analysis. For the chosen combination, the value of  $m_{ej}$  is the average between the two masses.

#### 4.3.1 Pre-selection

A sample of events enriched in SM background processes is selected to verify the background estimate in the  $eejj$  channel. The  $eejj$  preselection proceeds with the following kinematic cuts (all the reconstructed objects are required to pass the selection criteria described in Section 3):

- passing the signal triggers listed in Tables 8 and 9;
- exactly two electrons with  $p_T > 45$  GeV and  $|\eta| < 2.5$ ;
- at least 1 jet with  $p_T > 125$  GeV and  $|\eta| < 2.4$ ;
- at least 2 jets with  $p_T > 45$  GeV and  $|\eta| < 2.4$ ;
- $m_{ee} > 50$  GeV;
- $S_T > 300$  GeV;

At this stage of the selection, there is sufficient data to compare with the background predictions for all the observables employed in the final event selection. The descriptions of reconstructed quantities shown in the following figures are presented with this format:

- **the top plot** shows the distributions of the reconstructed quantity for data and background predictions described in Section 5. Signal samples are also shown for various leptoquark masses. The rightmost bin in the x-axis also includes the overflow bin.
- **the middle plot** shows the number of data events minus the number of background events (bin-by-bin), divided by the statistical uncertainty on this difference. It presents the agreement between data and background predictions in equivalent number of gaussian standard deviations.
- **the bottom plot** shows the bin-by-bin ratio between data and background prediction.

The distribution of the number of reconstructed primary vertices is shown in Figure 3. The  $p_T$  and  $\eta$  distributions of the two electrons and the two leading jets are shown in Figures 4, 5, 6, and 7. The  $m_{ee}$  distribution is shown in Figure 8. The  $S_T$  and  $m_{ej}$  distributions are shown in Figure 9.

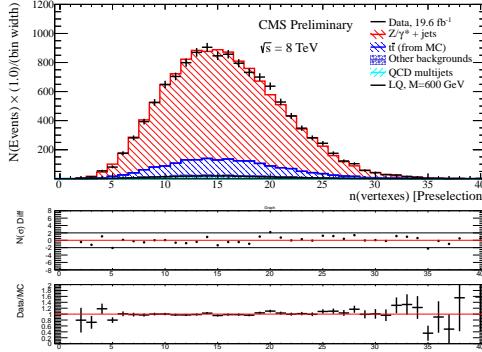


Figure 3: The distribution of the number of primary vertices for events passing the  $eejj$  preselection.

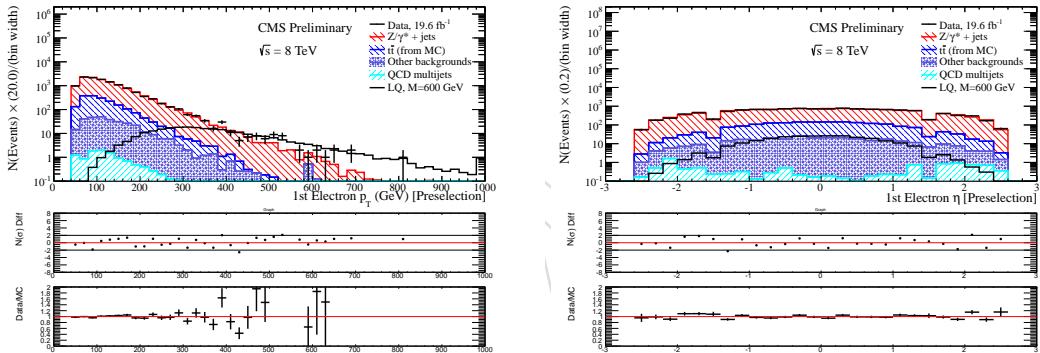


Figure 4: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $eejj$  preselection.

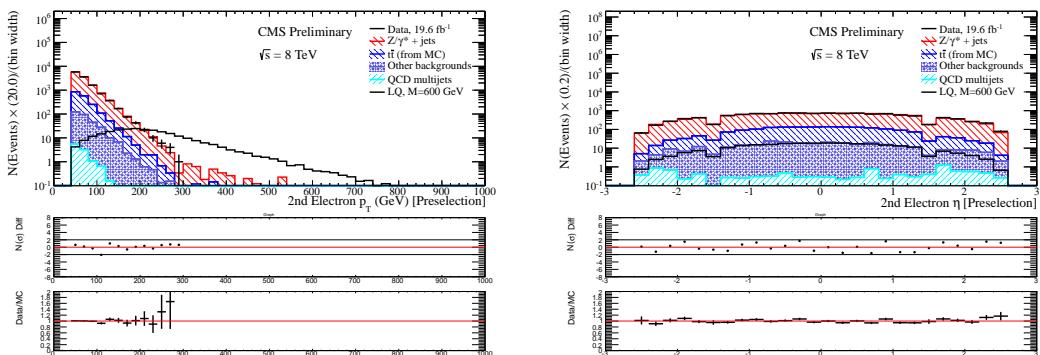


Figure 5: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) electron for events passing the  $eejj$  preselection.

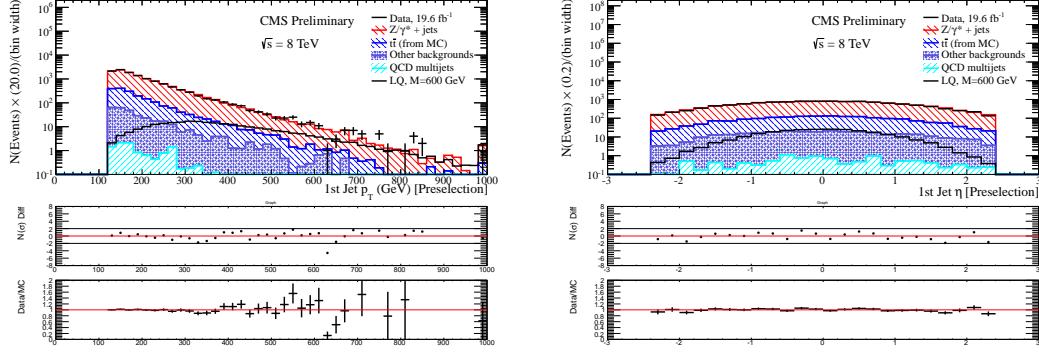


Figure 6: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $eejj$  preselection.

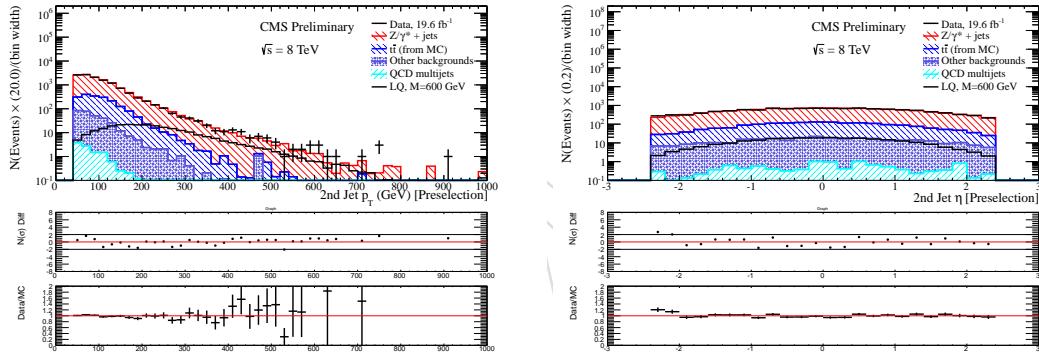


Figure 7: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $eejj$  preselection.

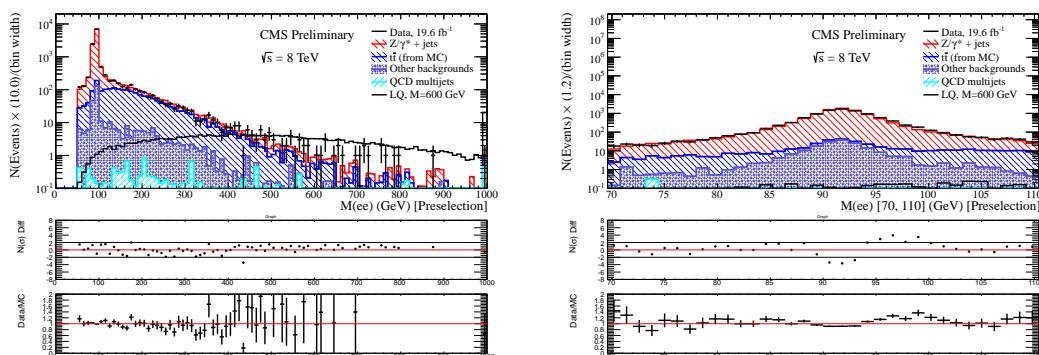


Figure 8: The  $m_{ee}$  distribution in the full range (left) and in a zoomed region around the  $Z^0$  mass peak (right) passing the  $eejj$  preselection.

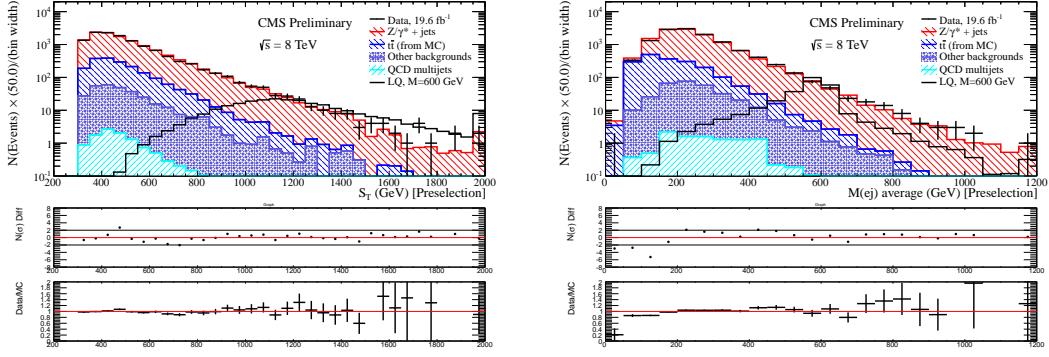


Figure 9: The  $S_T$ (left) and  $m_{ej}$ (right) distributions for events passing the  $eejj$  preselection.

### 4.3.2 Selection optimization

The final event selection criteria are optimized by maximizing the expected signal significance defined as  $S/\sqrt{S+B}$ , where  $S(B)$  is the expected number of signal (background) events passing the selection requirements. Three variables ( $S_T$ ,  $m_{ej}^{\min}$ , and  $m_{ee}$ ) are optimized simultaneously by scanning appropriate ranges of values. An optimized cut value is chosen for each variable at each leptoquark mass hypothesis under consideration. For each variable, the optimized cut values are then plotted as a function of the leptoquark mass hypothesis for which they have been chosen, and that plot is fit with a degree 2 polynomial within the leptoquark mass range of 300 GeV to 1000 GeV. The full leptoquark mass range is not used for the fit, due to the limited standard model background statistics available at cuts optimized for high leptoquark mass searches. These plots and fits are shown in Figure 10.

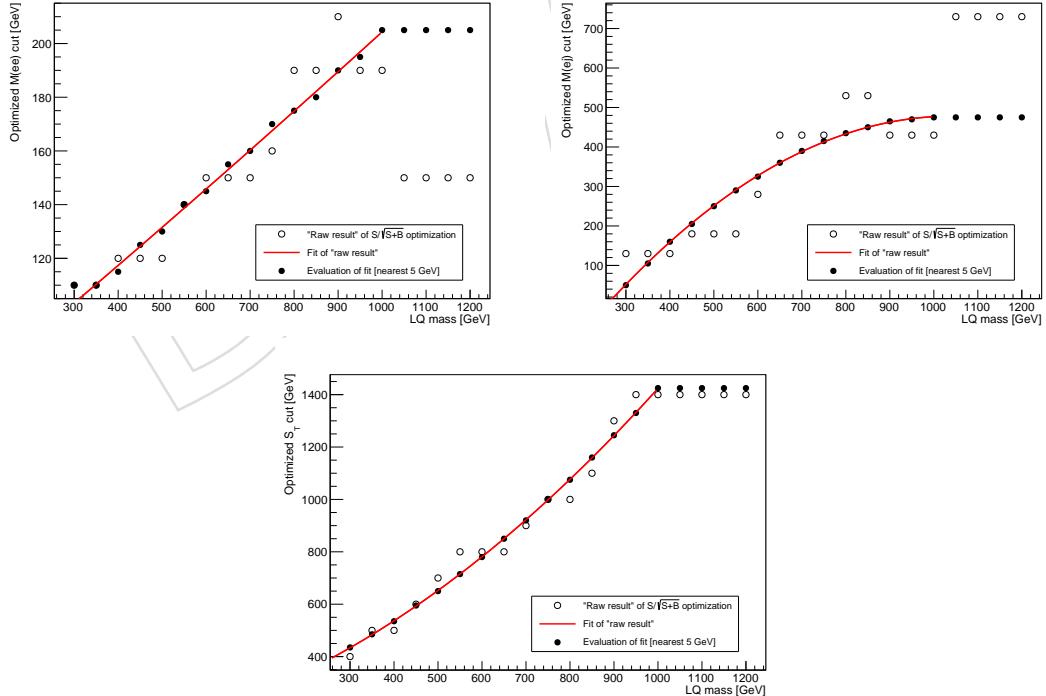


Figure 10: Optimized final selection criteria for the  $eejj$  analysis.

For leptoquark masses between 300 GeV and 1000 GeV, evaluations of these fits rounded to

301 the nearest 5 GeV are used as the final selection for the analysis. For leptoquark masses greater  
 302 than 1000 GeV, the cuts chosen for 1000 GeV leptoquarks are used. The final values chosen for  
 303 each leptoquark mass are given in Table 10.

Table 10: Optimized final selection criteria for the  $eejj$  analysis for different LQ mass hypotheses

	LQ mass ( $eejj$ )														
	300	350	400	450	500	550	600	650	700	750	800	850	900	950	$\geq 1000$
$S_T$ [GeV]	435	485	535	595	650	715	780	850	920	1000	1075	1160	1245	1330	1425
$m_{ee}$ [GeV]	110	110	115	125	130	140	145	155	160	170	175	180	190	195	205
$m_{ej}^{\min}$ [GeV]	50	105	160	205	250	290	325	360	390	415	435	450	465	470	475

### 304 4.3.3 Final selection

305 Table 11 shows the number of events for the data, the backgrounds, and the LQ signal, after ap-  
 306 plying the final, optimized  $eejj$  selection criteria summarized in Table 10. An excess is observed  
 307 for all levels of final selection optimized for  $M_{LQ} > 300$  GeV. This excess is most significant at  
 308  $M_{LQ} = 650$  GeV. Figures 11 and 12 show the distributions of  $S_T$  and the selected electron-jet  
 309 invariant mass,  $m_{ej}$ , after the full selection optimized for  $M_{LQ} = 450$  and 650 GeV, respectively.  
 310 These plots show that the excess is background-like (it does not peak in  $m_{ej}$ ). The dominant  
 311 background contributions are from  $t\bar{t}$  and  $Z+jets$  events, while the contribution from the other  
 312 backgrounds is below 10% for LQ masses within the current reach of this analysis.

313 The excess is studied in further detail in various appendices at the end of this note. The studies  
 314 in these appendices show the following:

- 315 • The events in data passing the final selection  $M_{LQ} = 650$  GeV were produced at a  
 316 relatively consistent rate through the 2012 data taking period (Appendix

Table 11: Number of events after final  $eejj$  selection. Only statistical errors are reported, except in the “Total Background” column, where systematic uncertainties are also reported.

$M_{LQ}$	LQ Signal	Z+Jets	$t\bar{t}$ (from data)	QCD (from data)	Other	Data	Total BG
Presel	-	$10538.4 \pm 35.8$	$1566.6 \pm 29.2$	$10.87 \pm 0.10$	$303.8 \pm 7.4$	12442	$12419.6 \pm 46.8$
300	$13560.2 \pm 80.1$	$462.2 \pm 7.4$	$724.3 \pm 19.8$	$5.282 \pm 0.052$	$62.1 \pm 4.6$	1244	$1253.94 \pm 21.67 \pm 44.58$ (syst)
350	$6473.9 \pm 33.3$	$332.1 \pm 6.2$	$352.0 \pm 13.8$	$3.215 \pm 0.036$	$37.7 \pm 3.6$	736	$725.10 \pm 15.57 \pm 31.03$ (syst)
400	$3089.3 \pm 15.0$	$203.2 \pm 4.8$	$153.7 \pm 9.1$	$1.696 \pm 0.023$	$23.8 \pm 2.9$	389	$382.40 \pm 10.72 \pm 18.75$ (syst)
450	$1508.1 \pm 7.2$	$112.9 \pm 3.5$	$86.9 \pm 6.9$	$0.890 \pm 0.016$	$11.8 \pm 2.0$	233	$212.44 \pm 7.99 \pm 10.41$ (syst)
500	$767.4 \pm 3.6$	$66.5 \pm 2.7$	$47.2 \pm 5.1$	$0.485 \pm 0.011$	$7.4 \pm 1.6$	148	$121.61 \pm 5.96 \pm 6.13$ (syst)
550	$410.5 \pm 1.9$	$37.4 \pm 2.1$	$25.8 \pm 3.7$	$0.2758 \pm 0.0084$	$3.7 \pm 1.1$	81	$67.24 \pm 4.40 \pm 3.44$ (syst)
600	$225.7 \pm 1.0$	$22.2 \pm 1.6$	$14.2 \pm 2.8$	$0.1527 \pm 0.0065$	$3.12 \pm 1.00$	57	$39.66 \pm 3.35 \pm 2.05$ (syst)
650	$125.85 \pm 0.58$	$14.0 \pm 1.2$	$5.4 \pm 1.7$	$0.0760 \pm 0.0040$	$1.05 \pm 0.47$	36	$20.49 \pm 2.14 \pm 1.28$ (syst)
700	$72.88 \pm 0.33$	$8.16 \pm 0.93$	$4.3 \pm 1.5$	$0.0448 \pm 0.0029$	$0.21 \pm 0.12$	17	$12.74 \pm 1.80 \pm 0.74$ (syst)
750	$43.10 \pm 0.20$	$4.88 \pm 0.69$	$1.55 \pm 0.90$	$0.0258 \pm 0.0023$	$0.078 \pm 0.038$	12	$6.53 \pm 1.13 \pm 0.44$ (syst)
800	$26.17 \pm 0.12$	$2.93 \pm 0.52$	$1.04 \pm 0.73$	$0.0193 \pm 0.0022$	$0.078 \pm 0.038$	7	$4.06 \pm 0.90 \pm 0.27$ (syst)
850	$15.978 \pm 0.072$	$2.34 \pm 0.48$	$0.52 \pm 0.52$	$0.0111 \pm 0.0015$	$0.042 \pm 0.028$	5	$2.91 \pm 0.71 \pm 0.21$ (syst)
900	$9.813 \pm 0.044$	$1.23 \pm 0.36$	$0.52 \pm 0.52$	$0.0069 \pm 0.0012$	$0.022 \pm 0.020$	3	$1.77 \pm 0.63 \pm 0.11$ (syst)
950	$6.086 \pm 0.028$	$0.89 \pm 0.29$	$0.00000^{+1.14000}_{-0.000}$	$0.00451 \pm 0.00085$	$0.022 \pm 0.020$	1	$0.912^{+1.178}_{-0.295} \pm 0.05$ (syst)
1000	$3.860 \pm 0.018$	$0.56 \pm 0.22$	$0.00000^{+1.14000}_{-0.000}$	$0.00374 \pm 0.00082$	$0.0025 \pm 0.0025$	1	$0.567^{+1.162}_{-0.223} \pm 0.05$ (syst)
1050	$2.576 \pm 0.011$	$0.56 \pm 0.22$	$0.00000^{+1.14000}_{-0.000}$	$0.00374 \pm 0.00082$	$0.0025 \pm 0.0025$	1	$0.567^{+1.162}_{-0.223} \pm 0.05$ (syst)
1100	$1.6936 \pm 0.0072$	$0.56 \pm 0.22$	$0.00000^{+1.14000}_{-0.000}$	$0.00374 \pm 0.00082$	$0.0025 \pm 0.0025$	1	$0.567^{+1.162}_{-0.223} \pm 0.05$ (syst)
1150	$1.1272 \pm 0.0047$	$0.56 \pm 0.22$	$0.00000^{+1.14000}_{-0.000}$	$0.00374 \pm 0.00082$	$0.0025 \pm 0.0025$	1	$0.567^{+1.162}_{-0.223} \pm 0.05$ (syst)
1200	$0.7498 \pm 0.0030$	$0.56 \pm 0.22$	$0.00000^{+1.14000}_{-0.000}$	$0.00374 \pm 0.00082$	$0.0025 \pm 0.0025$	1	$0.567^{+1.162}_{-0.223} \pm 0.05$ (syst)

### 317 4.4 Event Selection in the $evjj$ Channel

318 This section describes the event selection in the search for leptoquarks in the decay  
 319 channel  $LQL\bar{Q} \rightarrow eq\nu\bar{q}'$ . The leading (in  $p_T$ ) electron and the two leading (in  $p_T$ ) jets

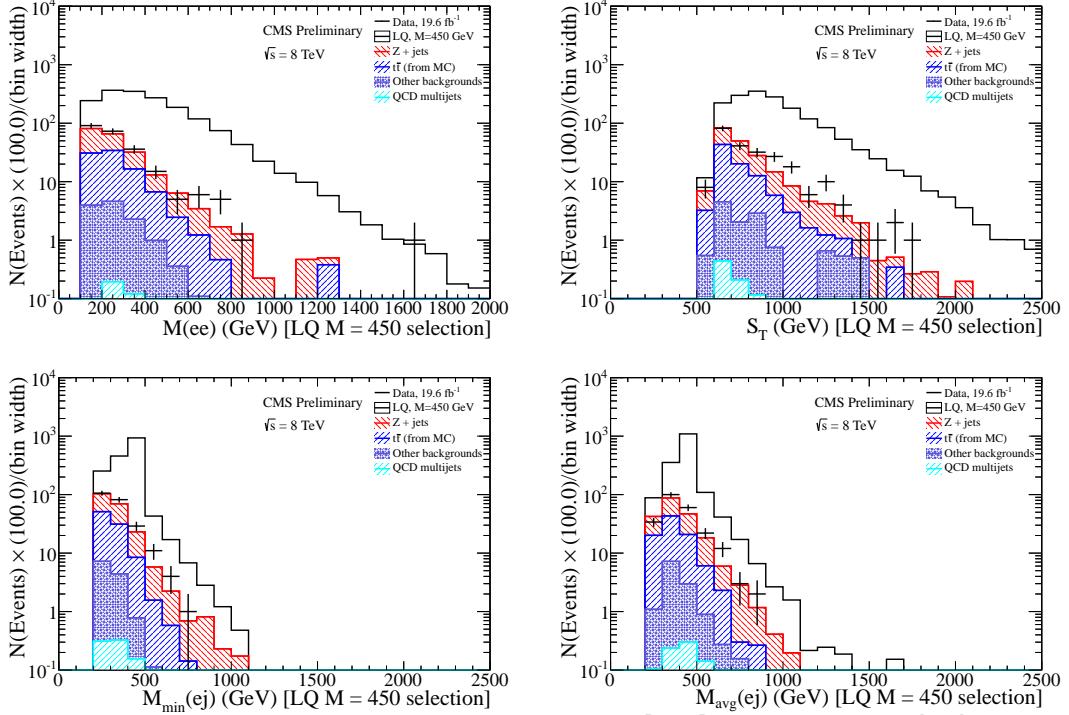


Figure 11: The  $m_{ee}$ (top left),  $S_T$ (top right),  $m_{ej}^{\min}$ (bottom left), and  $m_{ej}^{\text{avg}}$ (bottom right) distributions for events passing the full  $eejj$  selection optimized for  $M_{LQ} = 450$  GeV.

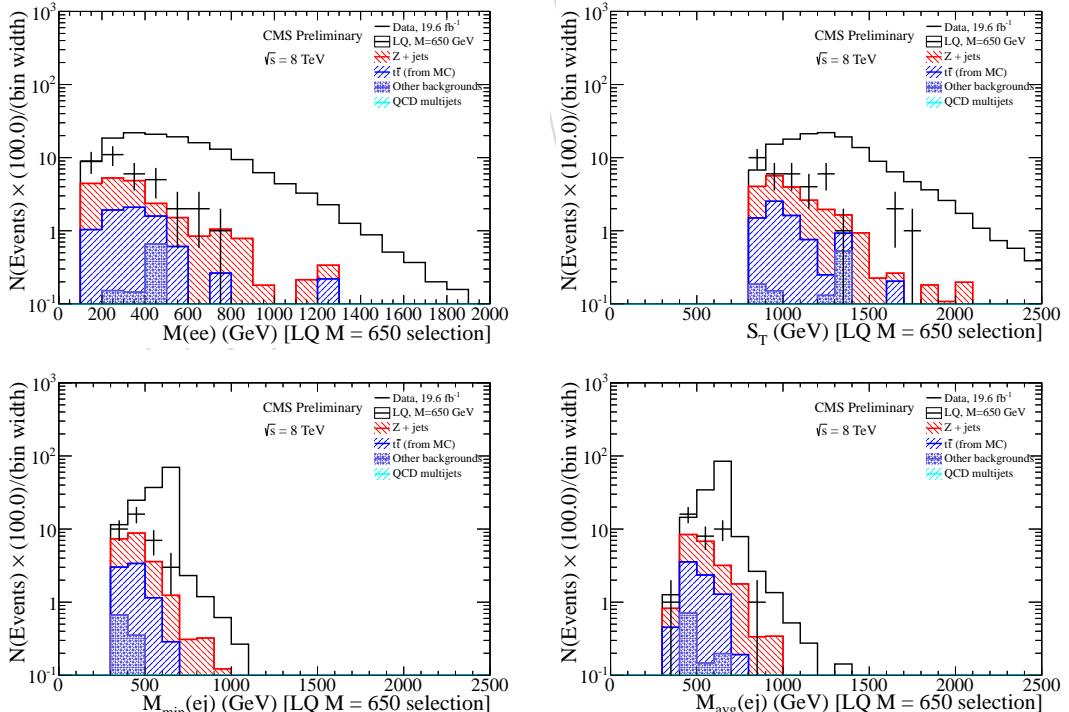


Figure 12: The  $m_{ee}$ (top left),  $S_T$ (top right),  $m_{ej}^{\min}$ (bottom left), and  $m_{ej}^{\text{avg}}$ (bottom right) distributions for events passing the full  $eejj$  selection optimized for  $M_{LQ} = 650$  GeV.

are used in this analysis. The definition of some reconstructed quantities that are used in the  $evjj$  analysis are described below:

- $\Delta\phi(E_T^{\text{miss}}, e)$ ,  $\Delta\phi(E_T^{\text{miss}}, j1)$ ,  $\Delta\phi(E_T^{\text{miss}}, j2)$ , the absolute values of the opening angles in  $\phi$  between the  $E_T^{\text{miss}}$  vector and the electron, the leading jet, and the second leading jet.
- $N_\mu$ , number of muons with  $p_T > 10$  and  $|\eta| < 2.4$ ;
- $\text{min}\Delta R(e, \text{jets})$  the minimum value of  $\Delta R$  between the electron and the two leading jets;
- $m_{T, ev} = \sqrt{2p_{T,e} \cdot E_T^{\text{miss}}(1 - \cos(\Delta\phi(E_T^{\text{miss}}, e)))}$ , the electron-neutrino transverse mass;
- $S_T$ , the scalar sum of the  $p_T$  of the electron, the  $E_T^{\text{miss}}$ , and the two leading jets;
- $m_{ej}$  and  $m_{T,\nu j}$ , the electron-jet invariant mass and the neutrino-jet transverse mass of the two leptoquark candidates, obtained from the electron, the  $E_T^{\text{miss}}$ , and the two leading jets. There are two possible ways to combine one electron, one neutrino (i.e.  $E_T^{\text{miss}}$ ) and two jets to form two leptoquark candidates: (e-j1, $\nu$ -j2) or (e-j2, $\nu$ -j1). The combination with the smaller difference between the electron-jet transverse mass and the neutrino-jet transverse mass is considered in this analysis.

#### 4.4.1 Pre-selection

A sample of events enriched in SM background processes is selected to verify the background estimate in the  $evjj$  channel. The  $evjj$  preselection proceeds with the following kinematic cuts (all of the reconstructed objects are required to pass the selection criteria described in Section 3):

- passing the signal triggers listed in Tables 8 and 9;
- exactly one electron,  $e$ , with  $p_T > 40$  GeV and  $|\eta| < 2.1$ ;
- at least 1 jet with  $p_T > 125$  GeV and  $|\eta| < 2.4$ ;
- at least 2 jets with  $p_T > 45$  GeV and  $|\eta| < 2.4$ ;
- $E_T^{\text{miss}} > 55$  GeV
- $\Delta\phi(E_T^{\text{miss}}, e) > 0.8$ ;
- $\Delta\phi(E_T^{\text{miss}}, j1) > 0.5$ ;
- $\text{min}\Delta R(e, \text{jets}) > 0.7$ ;
- $S_T > 300$  GeV

At this stage of the selection, there is sufficient data to compare with the background predictions for all the observables employed in the final event selection. The descriptions of reconstructed quantities shown in the following figures are presented with this format:

- **the top plot** shows the distributions of the reconstructed quantity for data and background predictions described in Section 5. Signal samples are also shown for various leptoquark masses. The rightmost bin in the x-axis also includes the overflow bin.
- **the middle plot** shows the number of data events minus the number of background events (bin-by-bin), divided by the statistical uncertainty on this difference. It presents the agreement between data and background predictions in equivalent number of gaussian standard deviations.

- 365     • the bottom plot shows the bin-by-bin ratio between data and background  
 366     prediction.

367     The distribution of the number of reconstructed primary vertices is shown in Figure  
 368     13. The  $p_T$  and  $\eta$  distributions of the electron and the two leading jets are shown in  
 369     Figures 14, 15, and 16. The  $E_T^{\text{miss}}$  distribution is shown in Figure 17. The  $\Delta\phi(E_T^{\text{miss}}, e)$ ,  
 370      $\Delta\phi(E_T^{\text{miss}}, j_1)$ ,  $\Delta\phi(E_T^{\text{miss}}, j_2)$ , and  $\text{min}\Delta R(e, \text{jets})$  distributions are shown in Figure 18.  
 371     The  $m_{T, ev}$  and  $S_T$  distributions are shown in Figure 19. The  $m_{ej}$  and  $m_{T\nu j}$  distributions  
 372     are shown in Figure 20.

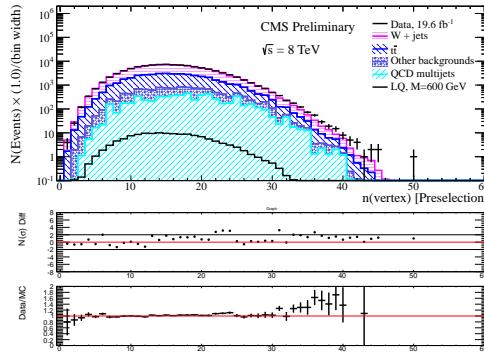


Figure 13: The distribution of the number of primary vertices for events passing the  $evjj$  preselection.

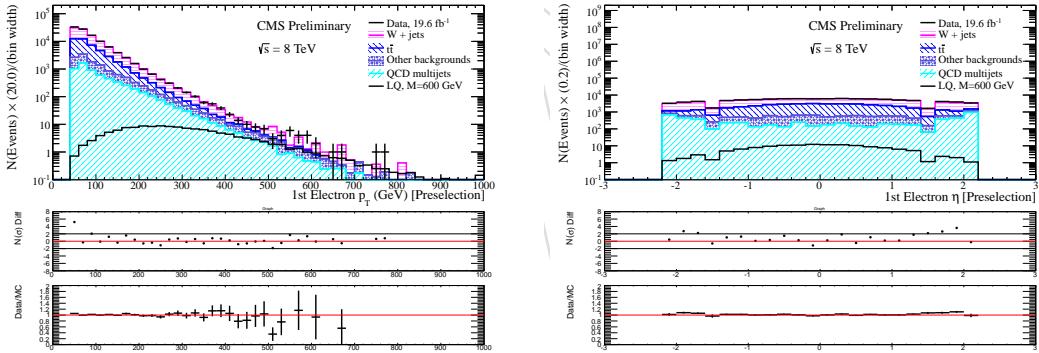


Figure 14: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $evjj$  preselection.

#### 4.4.2 Selection optimization

373     The final event selection criteria are optimized by maximizing the expected sig-  
 374     nificant signal defined as  $S/\sqrt{S+B}$ , where  $S(B)$  is the expected number of sig-  
 375     nificant (background) events passing the selection requirements. Four variables ( $S_T$ ,  $m_{ej}$ ,  
 376      $m_{T, ev}$ , and  $E_T^{\text{miss}}$ ) are optimized simultaneously by scanning appropriate ranges of  
 377     values. The fourth variable,  $E_T^{\text{miss}}$ , is added explicitly to remove QCD background.  
 378     An optimized cut value is chosen for each variable at each leptoquark mass hypothe-  
 379     sis under consideration. For each variable, the optimized cut values are then plotted  
 380     as a function of the leptoquark mass hypothesis for which they have been chosen,  
 381     and that plot is fit with a degree 2 polynomial within the leptoquark mass range of  
 382     300 GeV to 950 GeV. The full leptoquark mass range is not used for the fit, due  
 383     to the limited standard model background statistics available at cuts optimized for  
 384     high leptoquark mass searches. These plots and fits are shown in Figure 21.  
 385

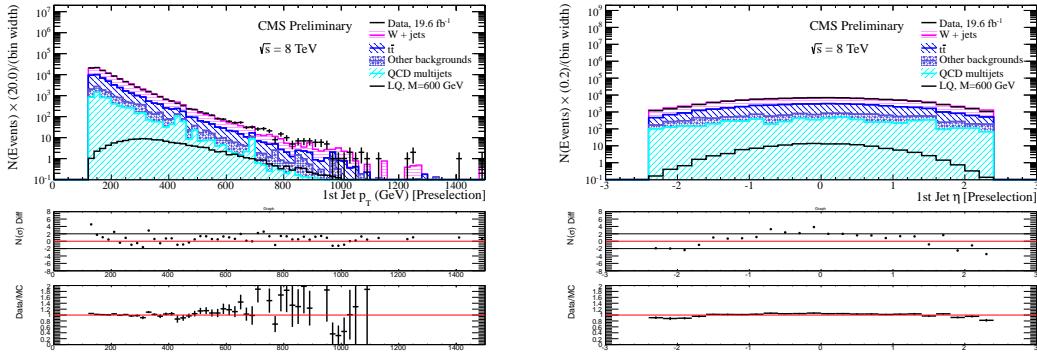


Figure 15: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $evjj$  preselection.

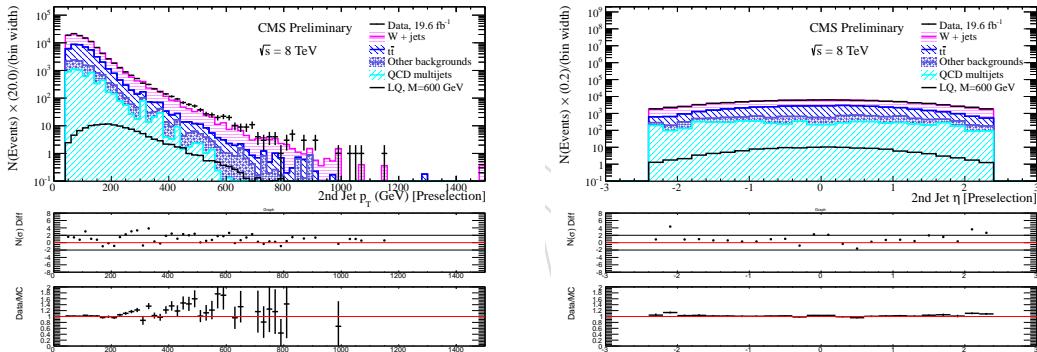


Figure 16: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $evjj$  preselection.

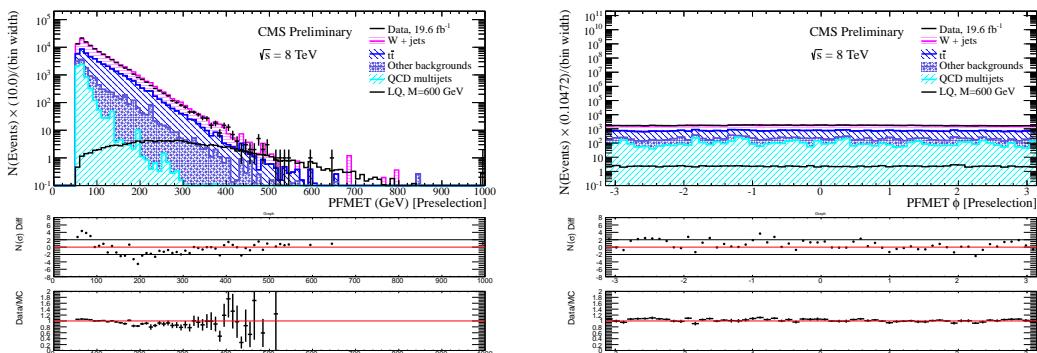


Figure 17: The  $E_T^{\text{miss}}$  (left) and  $\phi(E_T^{\text{miss}})$  (right) distributions for events passing the  $evjj$  preselection.

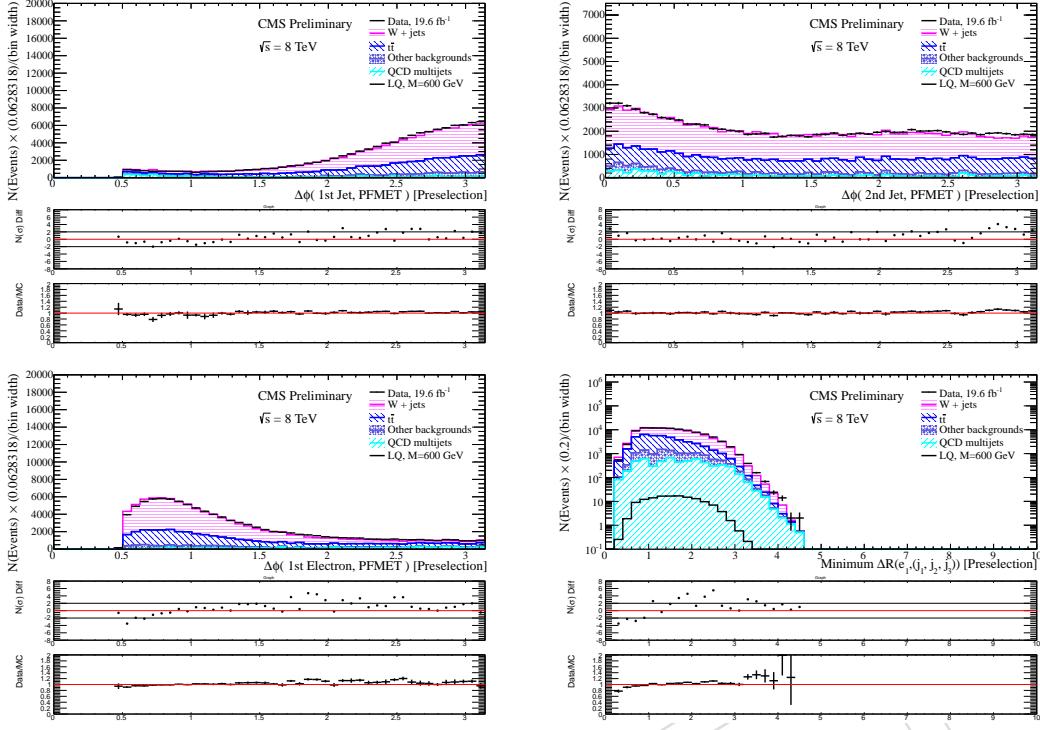


Figure 18: The distribution of  $\Delta|\phi|(E_T^{\text{miss}}, j_1)$  (top left),  $\Delta|\phi|(E_T^{\text{miss}}, j_2)$  (top right),  $\Delta|\phi|(E_T^{\text{miss}}, e)$  (bottom left), and  $\text{min}\Delta R(e, \text{jets})$  (bottom right) for events passing the  $e\nu jj$  preselection.

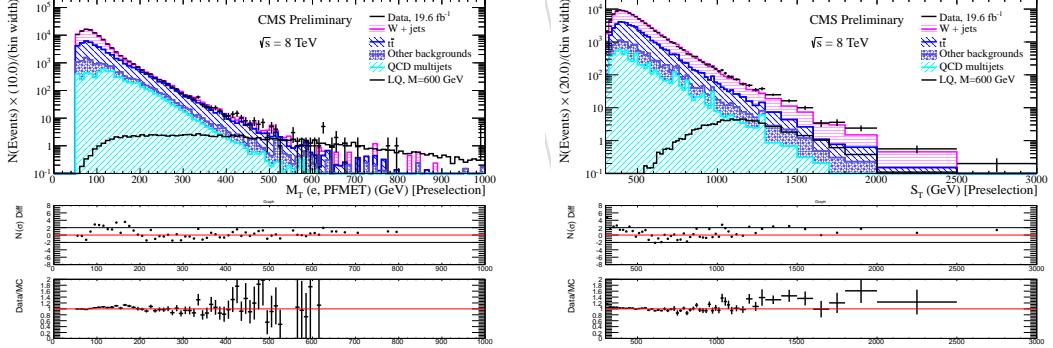


Figure 19: The  $m_{T, ee}$  (left) and  $S_T$  (right) distributions for events passing the  $e\nu jj$  preselection.

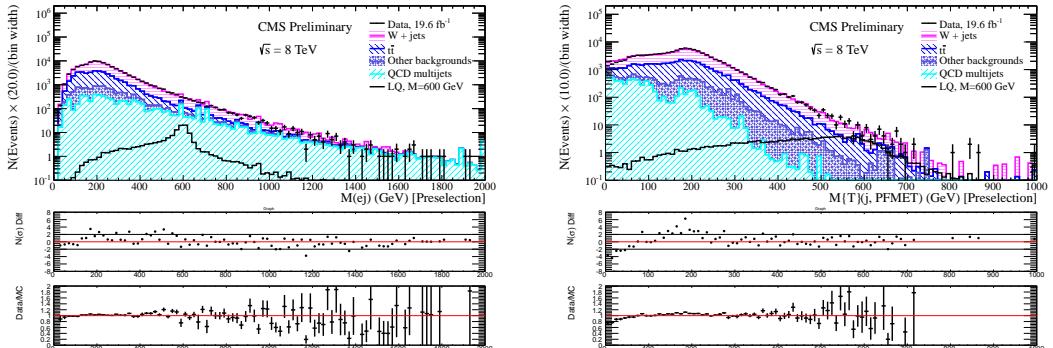
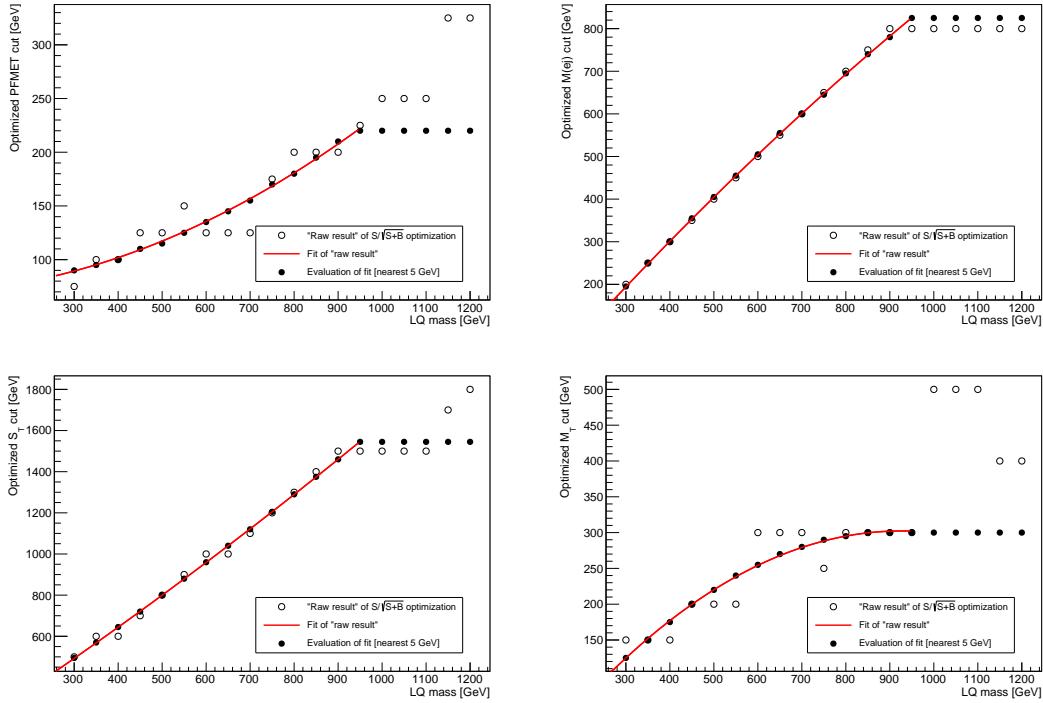


Figure 20: The  $m_{e,j}$  (left) and  $m_{T,v,j}$  (right) distributions for events passing the  $e\nu jj$  preselection.

Figure 21: Optimized final selection criteria for the  $evjj$  analysis.

Evaluations of these fits rounded to the nearest 5 GeV are used as the final selection for the analysis. The final values chosen for each leptoquark mass are given in Table 12.

Table 12: Optimized final selection criteria for the  $evjj$  analysis for different LQ mass hypotheses

	LQ Mass ( $evjj$ )													
	300	350	400	450	500	550	600	650	700	750	800	850	900	$\geq 950$
$S_T$ [GeV]	495	570	645	720	800	880	960	1040	1120	1205	1290	1375	1460	1545
PFMET [GeV]	90	95	100	110	115	125	135	145	155	170	180	195	210	220
$M(e)$ [GeV]	195	250	300	355	405	455	505	555	600	645	695	740	780	825
$M_T$ [GeV]	125	150	175	200	220	240	255	270	280	290	295	300	300	300

#### 4.4.3 Final selection

Table 13 shows the number of events for the data, the backgrounds, and the LQ signal, after applying the final, optimized  $eejj$  selection criteria summarized in Table 12. An excess is observed for all levels of final selection optimized for  $M_{LQ} = 300$  GeV. This excess is most significant at  $M_{LQ} = 650$  GeV. Figures 22 and 23 show the distributions of  $S_T$  and the selected electron-jet invariant mass,  $m_{ej}$ , after the full selection optimized for  $M_{LQ} = 450$  and 650 GeV, respectively. These plots show that the excess is background-like (it does not peak in  $m_{ej}$ ). The dominant background contributions are from  $t\bar{t}$  and  $Z+jets$  events, while the contribution from the other backgrounds is below 10% for LQ masses within the current reach of this analysis.

## 5 Backgrounds

The dominant SM backgrounds in the  $eejj$  ( $evjj$ ) channel are  $Z^0+jets$  ( $W+jets$ ) and  $t\bar{t}$ , but there are also contributions from other processes as summarized below.

Table 13: Number of events after final  $evjj$  selection. Only statistical errors are reported, except in the “Total Background” column, where systematic uncertainties are also reported.

$M_{LQ}$	LQ Signal	W+Jets	$t\bar{t}$	QCD	Other	Data	Total BG
Presel	-	$58284.8 \pm 197.0$	$32196.7 \pm 69.8$	$5950.5 \pm 20.1$	$6590.8 \pm 231.6$	105164	$103022.8 \pm 312.6$
300	$4765.5 \pm 51.1$	$822.1 \pm 22.4$	$1191.3 \pm 12.0$	$117.9 \pm 1.5$	$210.5 \pm 7.7$	2455	$2341.90 \pm 26.58 \pm 151.25$ (syst)
350	$2168.4 \pm 21.6$	$275.9 \pm 14.5$	$441.4 \pm 7.2$	$59.11 \pm 0.97$	$102.1 \pm 5.4$	908	$878.55 \pm 17.08 \pm 54.83$ (syst)
400	$971.1 \pm 9.6$	$110.4 \pm 7.8$	$184.2 \pm 4.7$	$32.88 \pm 0.69$	$51.5 \pm 3.8$	413	$378.98 \pm 9.91 \pm 23.44$ (syst)
450	$469.7 \pm 4.6$	$53.1 \pm 5.8$	$74.7 \pm 3.0$	$14.13 \pm 0.42$	$25.7 \pm 2.7$	192	$167.64 \pm 7.06 \pm 10.35$ (syst)
500	$232.7 \pm 2.3$	$20.5 \pm 3.3$	$34.4 \pm 2.0$	$7.76 \pm 0.30$	$15.3 \pm 2.1$	83	$77.99 \pm 4.41 \pm 4.64$ (syst)
550	$121.4 \pm 1.2$	$8.6 \pm 1.8$	$14.9 \pm 1.4$	$3.89 \pm 0.21$	$7.8 \pm 1.6$	44	$35.24 \pm 2.76 \pm 2.08$ (syst)
600	$66.37 \pm 0.66$	$2.3 \pm 1.0$	$7.08 \pm 0.93$	$2.29 \pm 0.17$	$4.6 \pm 1.2$	28	$16.27 \pm 1.84 \pm 0.98$ (syst)
650	$37.22 \pm 0.37$	$0.41 \pm 0.29$	$3.82 \pm 0.70$	$1.18 \pm 0.12$	$2.13 \pm 0.92$	18	$7.54 \pm 1.20 \pm 0.49$ (syst)
700	$21.74 \pm 0.21$	$0.41 \pm 0.29$	$2.61 \pm 0.60$	$0.85 \pm 0.10$	$0.58 \pm 0.24$	6	$4.45 \pm 0.71 \pm 0.34$ (syst)
750	$12.90 \pm 0.13$	$0.00^{+0.94}_{-0.00}$	$1.75 \pm 0.47$	$0.514 \pm 0.091$	$0.27 \pm 0.15$	4	$2.535^{+1.062}_{-1.019} \pm 0.21$ (syst)
800	$7.610 \pm 0.075$	$0.00^{+0.94}_{-0.00}$	$1.10 \pm 0.37$	$0.317 \pm 0.067$	$0.27 \pm 0.15$	3	$1.696^{+0.404}_{-0.399} \pm 0.13$ (syst)
850	$4.713 \pm 0.046$	$0.00^{+0.94}_{-0.00}$	$0.90 \pm 0.34$	$0.117 \pm 0.029$	$0.140 \pm 0.087$	2	$1.153^{+0.353}_{-0.353} \pm 0.08$ (syst)
900	$2.929 \pm 0.028$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.076 \pm 0.024$	$0.084 \pm 0.069$	1	$0.530^{+0.962}_{-0.226} \pm 0.04$ (syst)
950	$1.839 \pm 0.018$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)
1000	$1.306 \pm 0.012$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)
1050	$0.9022 \pm 0.0076$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)
1100	$0.6225 \pm 0.0050$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)
1150	$0.4308 \pm 0.0032$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)
1200	$0.2971 \pm 0.0022$	$0.00^{+0.94}_{-0.00}$	$0.37 \pm 0.21$	$0.069 \pm 0.023$	$0.084 \pm 0.069$	1	$0.524^{+0.962}_{-0.226} \pm 0.04$ (syst)

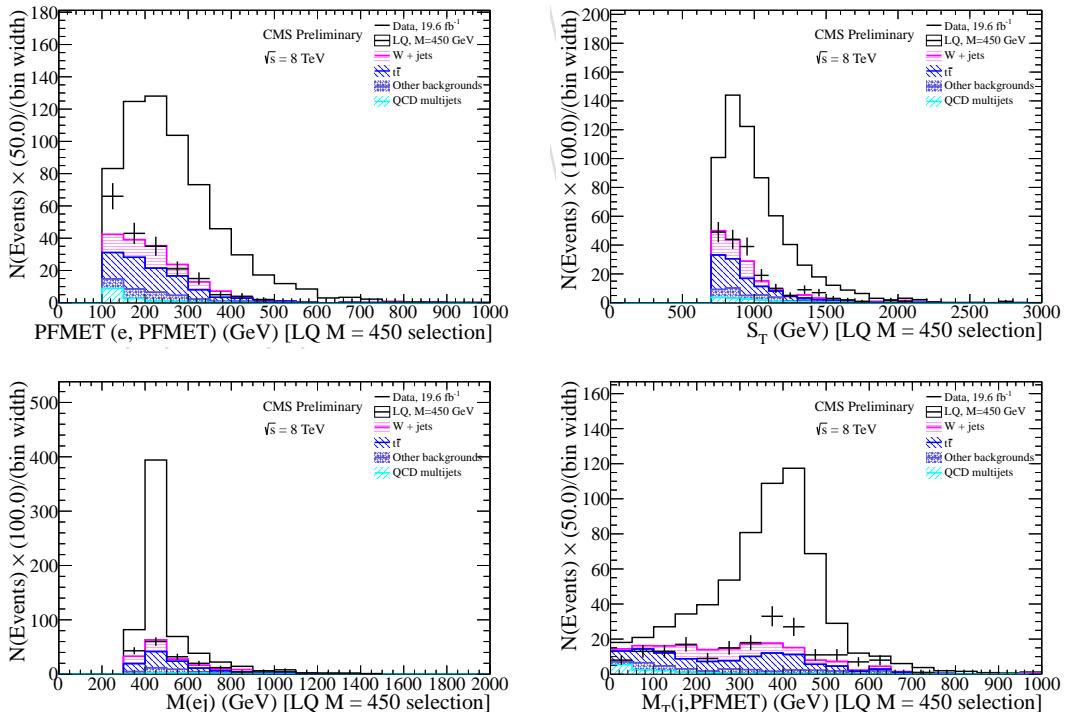


Figure 22: The  $E_T^{\text{miss}}$  (top left),  $S_T$ (top right),  $m_{ej}$ (bottom left), and  $m_{T,j\nu}$ (bottom right) distributions for events passing the full  $evjj$  selection optimized for  $M_{LQ} = 450$  GeV.

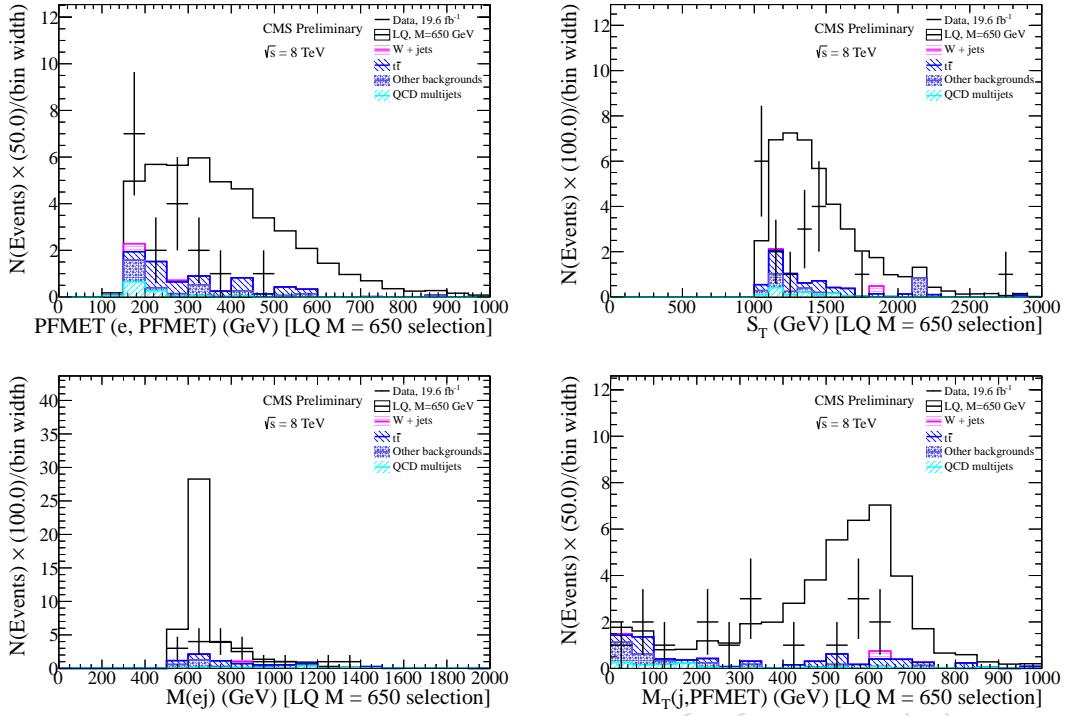


Figure 23: The  $E_T^{\text{miss}}$  (top left),  $S_T$  (top right),  $m_{\text{ej}}$  (bottom left), and  $m_{\text{T}, j_V}$  (bottom right) distributions for events passing the full  $\text{evjj}$  selection optimized for  $M_{\text{LQ}} = 650$  GeV.

The background estimation for the  $eejj$  analysis is as follows:

- the single-top,  $W+jets$ , diboson, and  $\gamma+jets$  backgrounds are taken from MC. More details are given in Section 5.1;
- the QCD multi-jet background is derived from data using the fake rate method described in Section 5.2;
- the  $t\bar{t}$  background is derived from data using an  $e\mu jj$  control sample. More details are given in Section 5.3.1;
- the  $Z^0+jets$  selection efficiency is derived from MC and the overall normalization of the sample is estimated from data at  $eejj$  preselection stage, as described in Section 5.4.1;

The background estimation for the  $evjj$  analysis is as follows:

- the single-top,  $Z^0+jets$ , diboson, and  $\gamma+jets$  backgrounds are taken from MC. More details are given in Section 5.1;
- the QCD multi-jet background is derived from data using the fake rate method described in Section 5.2;
- the  $t\bar{t}$  and  $W+jets$  selection efficiencies are derived from MC and the overall normalization of the two samples is estimated simultaneously from data, as described in Section 5.3.2;

The various steps listed above are described in order in the following sections.

## 5.1 Other Backgrounds

The following SM processes give a small contribution to the total background in both  $eejj$  and  $evjj$  analyses: single-top, diboson,  $\gamma+jets$ ,  $Z^0+jets$  (only in the  $evjj$  channel), and  $W+jets$  (only in the  $eejj$  channel). The selection efficiencies for these back-

Table 14: Loose identification criteria for gsf electrons needed for the QCD multijet background estimation. These criteria are used to select the loose  $eejj$  and  $evjj$  samples

Variable	Barrel threshold	Endcap threshold
$\sigma_{i\eta i\eta}$	$< 0.013$	$< 0.034$
$H/E$	$< 0.15$	$< 0.10$
$ d_{xy} $	$< 0.02$	$< 0.05$
Missing hits	$\leq 1$	$\leq 1$

425 grounds are derived from MC, and the samples are normalized to the cross sections  
 426 listed in Section 2.2.

## 5.2 QCD Multijet Background

### 5.2.1 Description of the method

429 The QCD multijet background in both the  $eejj$  and  $evjj$  channels is determined from  
 430 data using a fake rate method. Two data samples, “loose  $eejj$ ” and “loose  $evjj$ ”,  
 431 dominated by QCD multijet events are selected: the electrons are required to pass  
 432 “loose” identification requirements instead of the “tight” HEEP selection described  
 433 in Section 3.1, while the rest of the jet/ $E_T^{\text{miss}}$  requirements and kinematic selection  
 434 criteria remain unchanged.

435 The number of QCD multijet events in the  $eejj$  sample,  $N_{eejj}^{\text{QCD}}$ , at a given stage of the  
 436 selection is estimated by Equation 2.

$$N_{eejj}^{\text{QCD}} = \sum_{\text{loose } eejj \text{ events in data}} P(e_1, \text{tight} | e_1, \text{loose} : p_T, \eta) \cdot P(e_2, \text{tight} | e_2, \text{loose} : p_T, \eta) \quad (2)$$

437 Similarly, the number of QCD multijet events in the  $evjj$  sample,  $N_{evjj}^{\text{QCD}}$ , at a given  
 438 stage of the selection is estimated by Equation 3.

$$N_{evjj}^{\text{QCD}} = \sum_{\text{loose } evjj \text{ events in data}} P(e_{\text{tight}} | e_{\text{loose}} : p_T, \eta) \quad (3)$$

439 where  $e_{\text{loose}}$  is an electron passing the loose electron identification criteria described  
 440 in Table 14,  $e_{\text{tight}}$  is an electron passing the HEEP ID and isolation criteria described  
 441 in Section 3.1, and  $P(e_{\text{loose}} | e_{\text{tight}} : p_T, \eta)$  is the probability, or fake rate, that a loose  
 442 electron  $e_{\text{loose}}$  passes the HEEP ID and isolation requirements. The sum is performed  
 443 over the events in the loose  $eejj$  and  $evjj$  data samples passing the kinematic require-  
 444 ments at each considered stage of the event selection. The events in the loose  $eejj$  and  
 445  $evjj$  selection are selected online by prescaled single-photon triggers with different  
 446  $E_T$  thresholds (depending on the run range) as shown in Table 15. The lead “loose”  
 447 electron in the event is required to be matched to one of these triggers within a cone  
 448 of radius 0.5. Depending on the  $E_T$  of the triggered photon, each event passing the  
 449 loose  $eejj$  or  $evjj$  selection is reweighted in the sum of Equations 2 and 3 with a  
 450 weight equal to the prescale of the trigger with the highest threshold matched to the  
 451 “loose” electron in that event. The  $E_T$  threshold of the lowest single photon trigger  
 452 (`HLT_Photon30_CaloIdVL`) is 30 GeV; this value is 15 GeV below the electron  $p_T$   
 453 used in preselection for both analyses, enough to avoid any trigger threshold bias.

Table 15: Single-photon HLT paths used for the QCD multijet background estimation

HLT path	Run range	Effective $\mathcal{L}_{int}(\text{pb}^{-1})$
HLT_Photon30_CaloIdVL_v11	190456 - 190738	0.029672
HLT_Photon30_CaloIdVL_v12	190782 - 191419	0.086121
HLT_Photon30_CaloIdVL_v13	191691 - 196531	0.690924
HLT_Photon30_CaloIdVL_v14	198022 - 208686	2.043
HLT_Photon50_CaloIdVL_v7	190456 - 190738	0.231664
HLT_Photon50_CaloIdVL_v8	190782 - 191419	0.669828
HLT_Photon50_CaloIdVL_v9	191691 - 196531	5.374
HLT_Photon50_CaloIdVL_v10	198022 - 208686	15.894
HLT_Photon75_CaloIdVL_v10	190456 - 190738	1.385
HLT_Photon75_CaloIdVL_v11	190782 - 191419	4.019
HLT_Photon75_CaloIdVL_v12	191691 - 196531	32.243
HLT_Photon75_CaloIdVL_v13	198022 - 208686	95.363
HLT_Photon90_CaloIdVL_v7	190456 - 190738	2.769
HLT_Photon90_CaloIdVL_v8	190782 - 191419	8.038
HLT_Photon90_CaloIdVL_v9	191691 - 196531	69.509
HLT_Photon90_CaloIdVL_v10	198022 - 208686	198.024
HLT_Photon135_v4	190456 - 190738	96.404
HLT_Photon135_v5	190782 - 191419	398.151
HLT_Photon135_v6	191691 - 196531	543.603
HLT_Photon135_v7	198022 - 208686	12581
HLT_Photon150_v1	190456 - 190738	96.404
HLT_Photon150_v2	190782 - 191419	398.151
HLT_Photon150_v3	191691 - 196531	4824.
HLT_Photon150_v4	198022 - 208686	14304

454      **5.2.2 Fake rate determination**

455      The fake rate,  $P(e_{\text{loose}}|e_{\text{tight}} : p_T, \eta)$ , is determined as the ratio between the number  
 456      of HEEP electrons,  $N_{e,\text{tight}}$ , and the number of loose electrons  $N_{e,\text{loose}}$  identified with  
 457      the criteria listed in Table 14, in a sample of events in data passing the following  
 458      selection criteria:

- 459      1. The event must fire a single-photon trigger, as described in Table 15;
- 460      2. There must be exactly one loose electron, selected using the criteria described  
 461      in Table 14 having  $p_T > 10$  GeV;
- 462      3. There must be  $N_{\text{jet}}$  or more jets with  $p_T > 45$  GeV, with  $N_{\text{jet}} \geq 0, 1, 2$ , or 3;
- 463      4.  $\Delta R(e_{\text{loose}}, \text{jets}) > 0.7$ ;

464      The fake rates are calculated for electrons in the ECAL barrel and for two separate  
 465      ECAL endcap regions ( $|\eta| < 2.0$  and  $|\eta| > 2.0$ ). In addition, fake rates are calculated  
 466      independently for three separate jet multiplicity requirements, as described in cut  
 467      number 3, above. The contamination of real electrons in both  $N_{e,\text{tight}}$  and  $N_{e,\text{loose}}$   
 468      is subtracted from the data using MC predictions (the MC samples are normalized  
 469      using the cross sections listed in Section 2.2). The MC-corrected fake rate for the  
 470       $N_{\text{jet}} \geq 2$  case is shown in Figure 24 as a function of the loose electron  $p_T$ , separately  
 471      for the barrel and the two endcap regions.

472      Following the approach of the  $Z' \rightarrow ee$  analysis [21], each fake rate is fit with one  
 473      first-degree polynomial in the low- $p_T$  region and with different first-degree polyno-  
 474      mial in the high- $p_T$  region. The complete function is required to be continuous. The  
 475      fit results are reported in Table 16 for the four cases under consideration:  $N_{\text{jet}} \geq 0$ ,  
 476       $N_{\text{jet}} \geq 1$ ,  $N_{\text{jet}} \geq 2$ ,  $N_{\text{jet}} \geq 3$ . The fake rates that are used for the  $eejj$  and  $evjj$  QCD  
 477      multijet background estimation are from the  $N_{\text{jet}} \geq 2$  case.

478      **5.2.3 Estimation of the background**

479      At each step of the  $eejj$  and  $evjj$  selections, the shape of the kinematic distribu-  
 480      tions and the normalization of the QCD multijet background are estimated with the  
 481      method described in Section 5.2.1, using Equations 2 and 3. Tables 11 and 13 re-  
 482      port the QCD multijet contribution at different stages of the event selection for each  
 483      analysis using the  $N_{\text{jet}} \geq 2$  fake rates. After the final selection, the QCD multijet  
 484      contribution is  $\approx 1\%$  ( $\approx 3\%$ ) in the  $eejj$  ( $evjj$ ) analysis for leptoquark masses around  
 485      the current reach of this search.

486      **5.2.4 Closure test**

487      A closure test is performed in order to validate the QCD multijet background esti-  
 488      mation using the fake rate method. This closure test uses a control sample of QCD  
 489      multijet events (with a contamination from non-QCD processes of  $\approx 5\%$ ) selected by  
 490      the following requirements:

- 491      1. the event must fire a single photon trigger, as described in Table 15;
- 492      2. there must be exactly two loose electrons with  $p_T > 10$  GeV, selected using the  
 493      criteria described in Table 14;
- 494      3. both loose electrons must have  $p_T > 45$  GeV;
- 495      4. the lead loose electron must be matched to a single photon trigger;
- 496      5. there must be at least one jet with  $p_T > 45$  GeV.

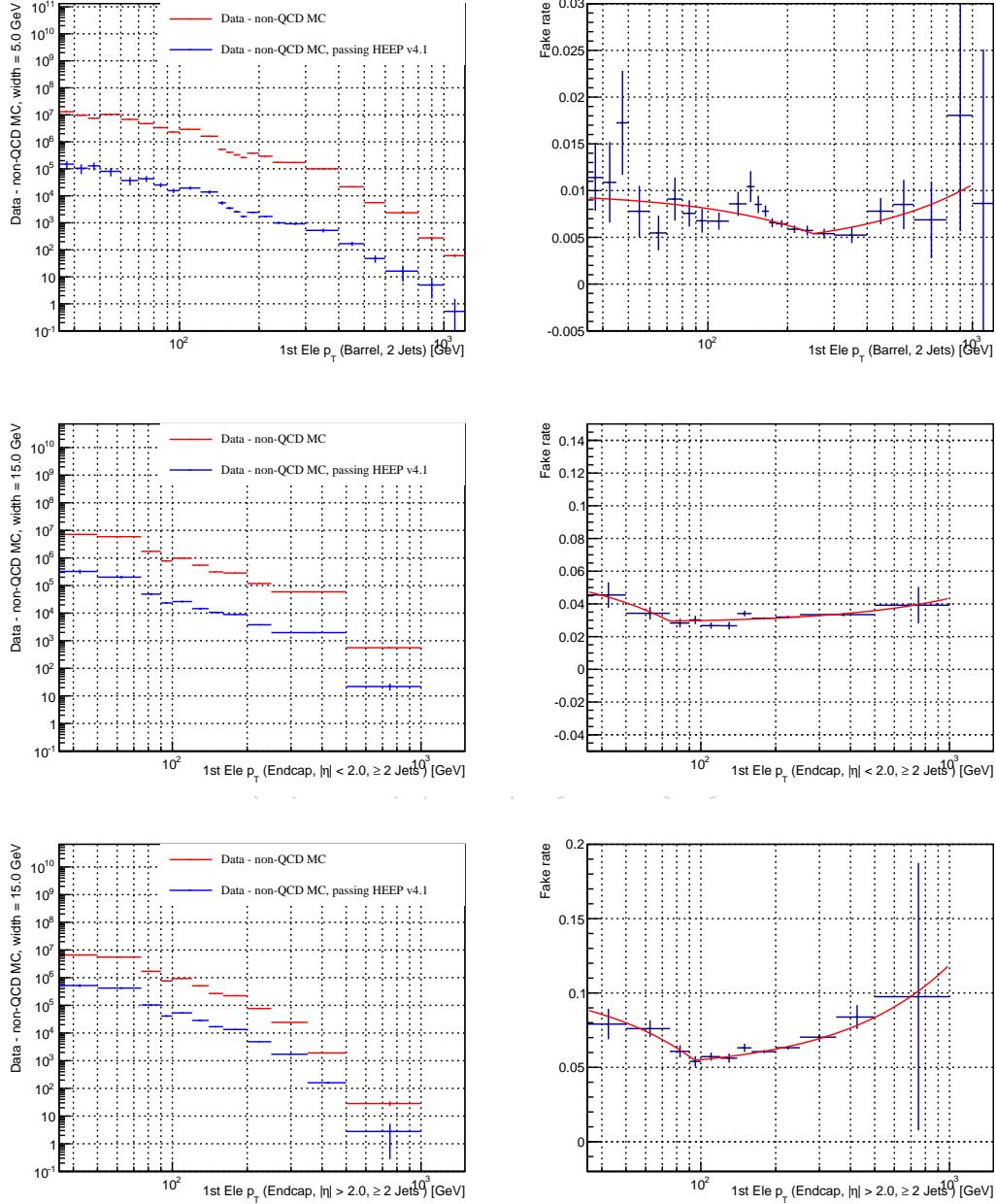


Figure 24: The  $p_T$  distribution of the “loose” and “tight” electrons in the fake rate sample (left) and the corresponding fitted fake rate vs. electron  $p_T$  (right) for selected events with  $N_{\text{jet}} \geq 2$ . The results are shown for the barrel (top row) the endcap with  $|\eta| < 2.0$  (middle row) and the endcap with  $|\eta| > 2.0$  (bottom row). The non-QCD contribution is subtracted from the electron  $p_T$  distributions.

Table 16: Fake rates and relative fit parameters

Jet multiplicity	Electron $\eta$ region	Fake rate
$N(\text{jet}) \geq 0$	$ \eta  < 1.442$	$\begin{cases} 1.562 \times 10^{-2} - 5.577 \times 10^{-5} \times E_T & \text{if } E_T < 182.7 \\ 5.581 \times 10^{-3} - 8.038 \times 10^{-7} \times E_T & \text{if } E_T \geq 182.7 \end{cases}$
	$1.56 <  \eta  < 2.0$	$\begin{cases} 8.684 \times 10^{-2} - 6.189 \times 10^{-4} \times E_T & \text{if } E_T < 87.6 \\ 3.310 \times 10^{-2} - 5.685 \times 10^{-6} \times E_T & \text{if } E_T \geq 87.6 \end{cases}$
	$2.0 <  \eta  < 2.5$	$\begin{cases} 1.064 \times 10^{-1} - 4.247 \times 10^{-4} \times E_T & \text{if } E_T < 106.9 \\ 5.788 \times 10^{-2} + 2.869 \times 10^{-5} \times E_T & \text{if } E_T \geq 106.9 \end{cases}$
$N(\text{jet}) \geq 1$	$ \eta  < 1.442$	$\begin{cases} 1.294 \times 10^{-2} - 3.900 \times 10^{-5} \times E_T & \text{if } E_T < 208.2 \\ 4.928 \times 10^{-3} - 5.071 \times 10^{-7} \times E_T & \text{if } E_T \geq 208.2 \end{cases}$
	$1.56 <  \eta  < 2.0$	$\begin{cases} 5.918 \times 10^{-2} - 2.964 \times 10^{-4} \times E_T & \text{if } E_T < 91.3 \\ 3.250 \times 10^{-2} - 4.138 \times 10^{-6} \times E_T & \text{if } E_T \geq 91.3 \end{cases}$
	$2.0 <  \eta  < 2.5$	$\begin{cases} 8.703 \times 10^{-2} - 2.213 \times 10^{-4} \times E_T & \text{if } E_T < 121.6 \\ 5.527 \times 10^{-2} + 3.982 \times 10^{-5} \times E_T & \text{if } E_T \geq 121.6 \end{cases}$
$N(\text{jet}) \geq 2$	$ \eta  < 1.442$	$\begin{cases} 9.855 \times 10^{-3} - 1.786 \times 10^{-5} \times E_T & \text{if } E_T < 250.0 \\ 3.641 \times 10^{-3} + 7.000 \times 10^{-6} \times E_T & \text{if } E_T \geq 250.0 \end{cases}$
	$1.56 <  \eta  < 2.0$	$\begin{cases} 6.331 \times 10^{-2} - 4.517 \times 10^{-4} \times E_T & \text{if } E_T < 75.0 \\ 2.830 \times 10^{-2} + 1.518 \times 10^{-5} \times E_T & \text{if } E_T \geq 75.0 \end{cases}$
	$2.0 <  \eta  < 2.5$	$\begin{cases} 1.083 \times 10^{-1} - 5.662 \times 10^{-4} \times E_T & \text{if } E_T < 94.5 \\ 4.805 \times 10^{-2} + 7.102 \times 10^{-5} \times E_T & \text{if } E_T \geq 94.5 \end{cases}$
$N(\text{jet}) \geq 3$	$ \eta  < 1.442$	$\begin{cases} 6.553 \times 10^{-3} - 2.828 \times 10^{-12} \times E_T & \text{if } E_T < 150.0 \\ 7.230 \times 10^{-3} - 4.514 \times 10^{-6} \times E_T & \text{if } E_T \geq 150.0 \end{cases}$
	$1.56 <  \eta  < 2.0$	$\begin{cases} 2.963 \times 10^{-2} - 4.803 \times 10^{-12} \times E_T & \text{if } E_T < 75.0 \\ 2.864 \times 10^{-2} + 1.308 \times 10^{-5} \times E_T & \text{if } E_T \geq 75.0 \end{cases}$
	$2.0 <  \eta  < 2.5$	$\begin{cases} 9.257 \times 10^{-2} - 3.453 \times 10^{-4} \times E_T & \text{if } E_T < 113.6 \\ 4.397 \times 10^{-2} + 8.260 \times 10^{-5} \times E_T & \text{if } E_T \geq 113.6 \end{cases}$

- 497        6.  $\Delta R(e_{\text{loose}}, \text{jets}) > 0.7$ ;
- 498        7.  $M_{ee} > 110$  GeV, to reduce contamination from  $Z^0 \rightarrow ee$  events;
- 499        8.  $E_T^{\text{miss}} < 100$  GeV, to reduce contamination from  $t\bar{t}$ ;
- 500        9.  $S_T > 200$  GeV, where  $S_T$  is defined as the scalar sum of the two electrons and  
501        the leading (in  $p_T$ ) jet;

502        A prediction is made on how many of these events will have one electron passing  
503        the identification and isolation criteria described in Section 3.1,  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,pred.}}$  using  
504        the  $N_{\text{jet}} \geq 1$  fake rates calculated in Section 5.2.2 and following a similar procedure  
505        as the one described in the previous sections:

$$N_{\text{loose } e, \text{ tight } e}^{\text{QCD,pred.}} = \sum_{\text{loose } e, \text{ loose } e, j \text{ events}} P(e_1, \text{tight} | e_1, \text{loose} : p_T, \eta) + P(e_2, \text{tight} | e_2, \text{loose} : p_T, \eta) \quad (4)$$

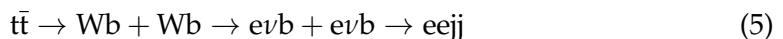
506        This prediction is then compared with the actual number of events in data where one  
507        of the two loose electrons also passes the tight selection criteria,  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,actual}}$ . In this  
508        comparison the contribution of  $\approx 55\%$  from non-QCD multijet processes contaminating  
509        the actual “loose  $e$ , tight  $e$ ,  $j$ ” data sample is subtracted using an estimation from  
510        MC (the samples are normalized using the cross sections listed in Section 2.2).

511        The predicted and observed values of “loose  $e$ , tight  $e$ ,  $j$ ” events corresponding to  
512         $19.6 \text{ fb}^{-1}$  of data are:  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,pred.}} = 13100 \pm 400$  and  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,actual}} = 12100 \pm 400$   
513        (the latter after MC subtraction). Statistical errors derived from the fake rate fit pa-  
514        rameters and the uncertainties on the MC predictions used in the subtraction are  
515        included. The ratio between the two values is  $1.08 \pm 0.05$ . If we apply an  $S_T^2$  cut  
516        of 435 GeV (corresponding to the  $S_T$  cut at the  $eejj$  final selection optimized for lep-  
517        toquarks with a mass of 300 GeV), the values become  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,pred.}} = 877 \pm 46.7$   
518        and  $N_{\text{loose } e, \text{ tight } e}^{\text{QCD,actual}} = 600 \pm 53$ , and their ratio is  $1.46 \pm 0.15$ . We quote therefore  
519        a 60% (30%) uncertainty on this method at final selection level for the  $eejj$  ( $evjj$ )  
520        analyses. Figure 25 shows the distributions of electron and jet reconstructed quan-  
521        tities for “loose  $e$ , tight  $e$ ,  $j$ ” events; the QCD multijet prediction using the fake rate  
522        method is compared with the actual data sample and an acceptable agreement is  
523        observed in both the shape and the normalization of the two samples within the  
524        quoted uncertainties.

### 5.3 $t\bar{t}$ Background

#### 5.3.1 $t\bar{t}$ Background in the $eejj$ channel

The  $t\bar{t}$  background in the  $eejj$  channel mainly comes from the process:



This contribution (both in terms of number of events and shape of kinematic dis-  
tributions) is estimated from a control sample containing one electron, one muon,  
and at least two jets ( $e\mu jj$  final state, where one of the Ws decays to an electron-  
neutrino pair and the other to a muon-neutrino pair). The  $e\mu jj$  sample is dominated  
by  $t\bar{t}$  events with a small contamination of  $\approx 2\%$ , mainly from diboson events, as

---

<sup>2</sup>defined as the scalar sum of the  $p_T$  values of the two loose electrons and the leading jet

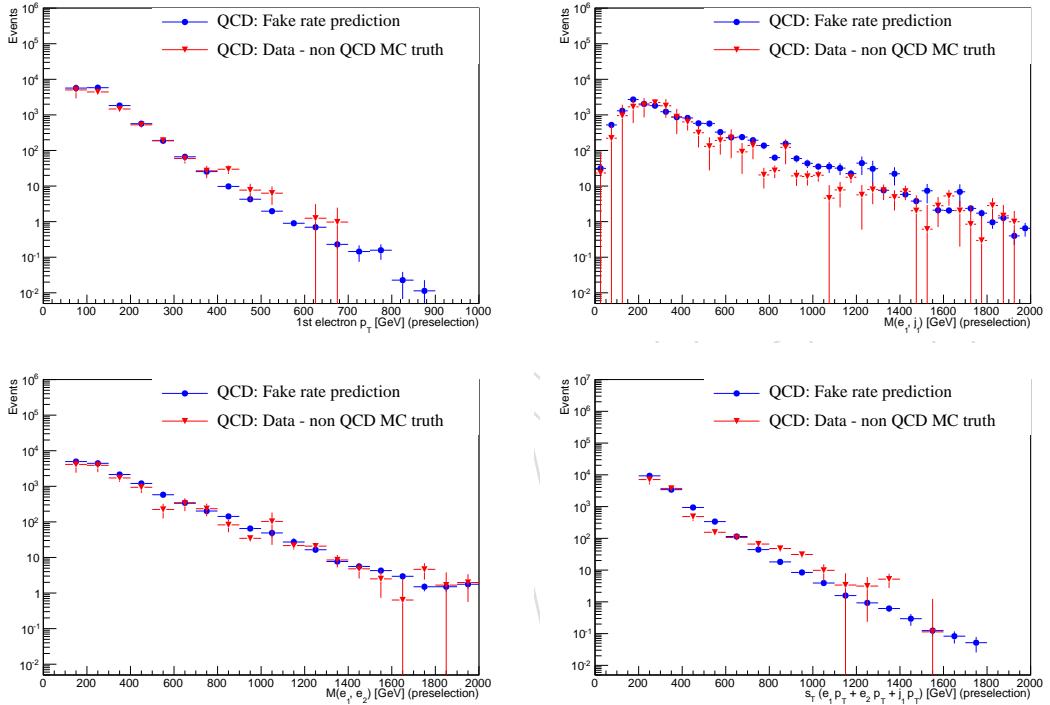


Figure 25: The leading electron  $p_T$  (top left),  $m_{e_j}$ (top right),  $m_{ee}$ (bottom left), and  $S_T$ (bottom right) distributions for the “loose  $e$ , tight  $e, j$ ” events. The QCD multijet prediction using the fake rate method is compared with the actual data sample (after the MC subtraction of non-QCD events for the latter sample).

derived from MC. The  $e\mu jj$  events are selected online using unprescaled muon triggers depending on the run period as reported in Table 17. The  $eejj$  preselection and final selection requirements described in Section 4.3 are applied to the  $e\mu jj$  sample as well, where the muon is treated as an electron. LQ events do not produce  $e\mu jj$  final states, under the assumption of no mixing between the generations of leptons and quarks. At generator level and before any event selection bias, the  $t\bar{t}$  background is expected to produce exactly twice as many  $e\mu jj$  events as  $eejj$  events with the same kinematic properties on average. In data, the relation between the number of  $eejj$  and  $e\mu jj$  events at any given stage of the selection must be corrected for trigger, reconstruction, ID, and isolation efficiencies of leptons:

$$N_{eejj}^{\text{data}} = \mathcal{C} \times N_{e\mu jj}^{\text{data}} = \frac{1}{2} \times \frac{\epsilon_{ee}^{\text{trigger}}}{\epsilon_{e\mu}^{\text{trigger}}} \times \frac{\epsilon_e^{\text{reco}/\text{ID}/\text{Iso}}}{\epsilon_\mu^{\text{reco}/\text{ID}/\text{Iso}}} \times N_{e\mu jj}^{\text{data}} \quad (6)$$

where :

- $\epsilon_{ee}^{\text{trigger}}$  ( $\epsilon_{e\mu}^{\text{trigger}}$ ) is the trigger efficiency to select di-electron (electron-muon) events with the HLT paths listed in Tables 8 and 9 (Table 17). The efficiency  $\epsilon_{ee}^{\text{trigger}}$  of the single electron plus dijet trigger on events with two HEEP electrons is greater than 99.8% and taken to be 100%. The efficiency  $\epsilon_{e\mu}^{\text{trigger}}$  is binned in muon  $|\eta|$  and is taken to be 0.94 (for  $|\eta| \leq 0.9$ ), 0.84 (for  $0.9 < |\eta| \leq 1.2$ ), 0.82 (for  $1.2 < |\eta| \leq 2.1$ ). To account for statistical uncertainties and variation over different data taking periods, a conservative uncertainty of 1% is placed on these efficiencies.
- $\epsilon_e^{\text{reco}/\text{ID}/\text{Iso}}$  ( $\epsilon_\mu^{\text{reco}/\text{ID}/\text{Iso}}$ ) is the single-electron (single-muon) reconstruction, ID, and isolation efficiency for trigger lepton candidates. The efficiency ratio is taken from  $t\bar{t}$  simulation at  $eejj$  preselection level:

$$\epsilon_e^{\text{reco}/\text{ID}/\text{Iso}} / \epsilon_\mu^{\text{reco}/\text{ID}/\text{Iso}} = 2 \times N_{eejj}^{\text{MC}} / N_{e\mu jj}^{\text{MC}} = 0.974 \pm 0.011 \text{ (stat)} \quad (7)$$

Table 17: Single-muon HLT paths used for the  $t\bar{t}$  background estimation in the  $eejj$  analysis

HLT path	Run range
HLT_Mu40_eta2p1_v9	190456 - 196531
HLT_Mu40_eta2p1_v10	198063 - 199608
HLT_Mu40_eta2p1_v11	199698 - 208686

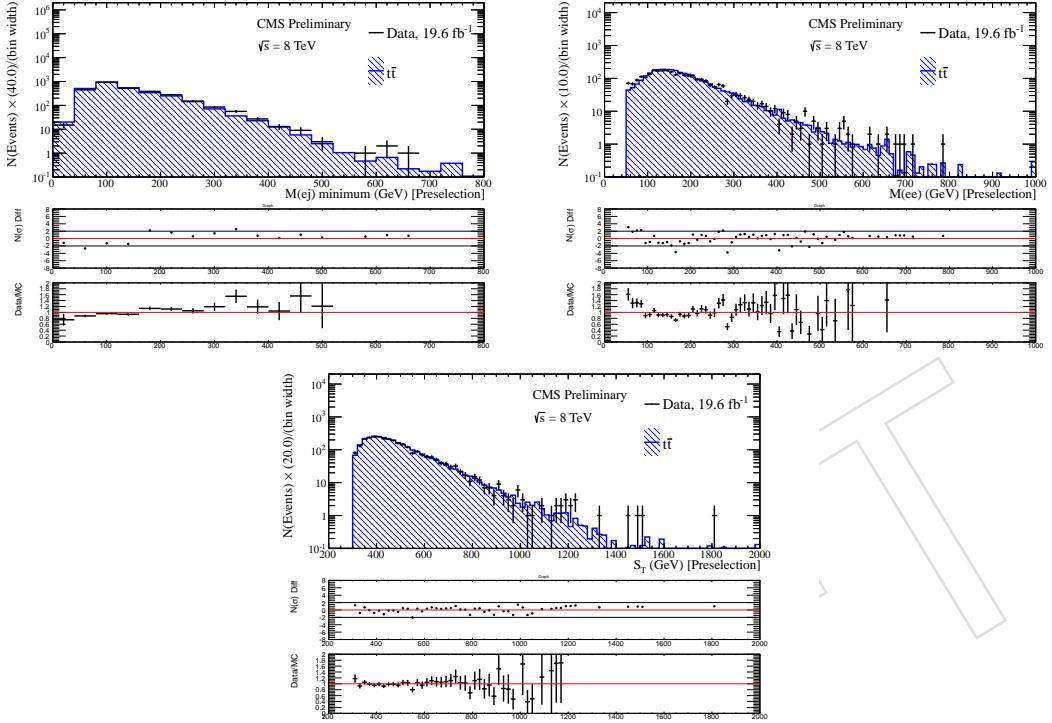
The  $t\bar{t}$  background prediction is thus obtained by weighting the  $e\mu jj$  events by the factors in table 18.

Table 18: Weights applied to events in the  $e\mu jj$  sample in order to estimate the contribution of  $eejj$  events from  $t\bar{t}$

Muon $ \eta $ range	Weight applied to $e\mu jj$ events
$0.0 <  \eta  \leq 0.9$	$\mathcal{C} = 0.458 \pm 0.005 \text{ (stat)} \pm 0.005 \text{ (syst)}$
$0.9 <  \eta  \leq 1.2$	$\mathcal{C} = 0.409 \pm 0.005 \text{ (stat)} \pm 0.005 \text{ (syst)}$
$1.2 <  \eta  \leq 2.1$	$\mathcal{C} = 0.400 \pm 0.005 \text{ (stat)} \pm 0.005 \text{ (syst)}$

A comparison can be made between the data-driven method described in this section and the MADGRAPH  $t\bar{t}$  Monte Carlo used elsewhere in this analysis at  $eejj$  preselection. At  $eejj$  preselection, the data-driven method predicts  $1579.6 \pm 29.3$   $t\bar{t}$  events,

541 while the  $eejj$  preselection sample of  $t\bar{t}$  Monte Carlo predicts  $1582.2 \pm 13.8$   $t\bar{t}$  events  
 542 (uncertainties are statistical only). In addition to raw event yield, good agreement is  
 543 observed between the data-driven method and the  $eejj$   $t\bar{t}$  Monte Carlo in modeling  
 544 the shape of the distributions used in the  $eejj$  final selection:  $m_{ej}^{\min}$ ,  $m_{ee}$ , and  $S_T$ . The  
 545 distributions for these variables at  $eejj$  preselection are shown in Figure 26.



561 Figure 26: Comparison between events from data in the  $eejj$  control sample (weighted ac-  
 562 cording to Table 18) and events from  $eejj$  Monte Carlo at preselection. The distributions pro-  
 563 vided correspond to the variables used at final selection in the  $eejj$  analysis:  $m_{ej}^{\min}$  (top left plot),  
 564  $m_{ee}$  (top right plot), and  $S_T$  (bottom plot).

### 566 5.3.2 $t\bar{t}$ (and W+jets) Background in the $evjj$ channel

567 The  $t\bar{t}$  and W+jets selection efficiencies in the  $evjj$  analysis are derived from MC. The  
 568  $t\bar{t}$  and W+Jets normalizations are obtained by comparing data and simulation after  
 569 two separate selection criteria: one that enriches the samples with the  $t\bar{t}$  events and  
 570 one that enriches the samples with the W+jets events.

571 The following selection (selection 1) enriches the samples with W+jets events, with  
 572 a contamination of 30%, mainly coming from  $t\bar{t}$  events.

- 573 •  $evjj$  preselection;
- 574 •  $50 < m_{T, ev} < 110$  GeV
- 575 • fewer than 4 jets with  $p_T > 40$  GeV and  $|\eta| < 2.4$

576 The following selection (selection 2) enriches the samples with  $t\bar{t}$  events, with a con-  
 577 tamination of 31%, mainly coming from W+jets and QCD multijet events.

- 578 •  $evjj$  preselection;
- 579 •  $50 < m_{T, ev} < 110$  GeV
- 580 • at least 4 jets with  $p_T > 40$  GeV and  $|\eta| < 2.4$

The results of the two selections can be used to form a system of equations:

$$\begin{cases} N_{\text{data}}^i = \mathcal{R}_{\bar{t}\bar{t}} N_{\bar{t}\bar{t}}^i + \mathcal{R}_W N_W^i + N_{\text{QCD}}^i + N_{\text{Others}}^i \\ N_{\text{data}}^2 = \mathcal{R}_{\bar{t}\bar{t}} N_{\bar{t}\bar{t}}^2 + \mathcal{R}_W N_W^2 + N_{\text{QCD}}^2 + N_{\text{Others}}^2 \end{cases} \quad (8)$$

where  $N_{\text{data}}^i$ ,  $N_W^i$ ,  $N_{\text{Others}}^i$ ,  $N_{\bar{t}\bar{t}}^i$ , and  $N_{\text{QCD}}^i$  are, respectively, the number of events in data, W+jets MC, other MC backgrounds (single-top, diboson, etc.),  $\bar{t}\bar{t}$  MC (before rescaling), and QCD multijet events obtained with the method described in Section 5.2, passing selection  $i$ . Solving the system yields the following rescaling factors for the MADGRAPH  $\bar{t}\bar{t}$  and W+jets samples:

$$\begin{aligned} \mathcal{R}_{\bar{t}\bar{t}} &= 0.97 \pm 0.02 \text{ (stat)} \pm 0.01 \text{ (syst)} \\ \mathcal{R}_W &= 0.85 \pm 0.01 \text{ (stat)} \pm 0.01 \text{ (syst)} \end{aligned} \quad (9)$$

For this study, all the MC samples are normalized using the cross sections listed in Section 2.2. These factors are already included in the  $\bar{t}\bar{t}$  and W+jets MC predictions for the  $eejj$  analysis shown in this note. Figure 27 shows the distributions of several reconstructed quantities for events passing the criteria in selection 2 listed above. A good agreement between data and background and prediction is found in the shape of all of these distributions.

## 5.4 V+jets Background

### 5.4.1 $Z^0$ +jets Background in the $eejj$ channel

The  $Z^0$ +jets selection efficiency in the  $eejj$  analysis is derived from MC. The  $Z^0$ +jets MC sample normalization is obtained by comparing data and simulation as described below.

After applying the  $eejj$  preselection and requiring that the invariant mass of the di-electron pair is close to the  $Z^0$  peak,  $70 < m_{ee} < 110$  GeV, the sample is dominated by  $Z^0$ +jets events with a contamination from other SM processes of  $\approx 4\%$ , estimated from MC, mostly coming from  $\bar{t}\bar{t}$  and diboson events. At this stage of the selection, the data is compared to the background predictions and an overall good agreement is found in the shape of the distributions of all of the observables employed in the final selection. A data/MC rescaling factor is then computed for the  $Z^0$ +jets background:

$$\mathcal{R}_{Z^0} = \frac{N_{\text{data}} - (N_{\text{Others}} + N_{\text{QCD}})}{N_{Z^0}} = 0.97 \pm 0.01 \text{ (stat)} \pm 0.00004 \text{ (syst)} \quad (10)$$

where where  $N_{\text{data}}^i$ ,  $N_Z^i$ ,  $N_{\text{Others}}^i$ ,  $N_{\text{QCD}}^i$  are, respectively, the number of events in data,  $Z$ +jets MC, other MC backgrounds (single-top, diboson, etc.), and QCD multijet events obtained with the method described in Section 5.2. This factor is already included in the  $Z^0$ +jets MC predictions for the  $eejj$  analysis shown in this note. The systematic uncertainty on the rescaling factor comes from the systematic uncertainty on the data-driven QCD and  $\bar{t}\bar{t}$  estimates.

### 5.4.2 W+jets Background in the $evjj$ channel

The W+jets selection efficiency in the  $evjj$  analysis is derived from MC. The W+jets MC sample normalization is obtained by comparing data and simulation, simultaneously with the  $\bar{t}\bar{t}$  background, as described in Section 5.3.2.

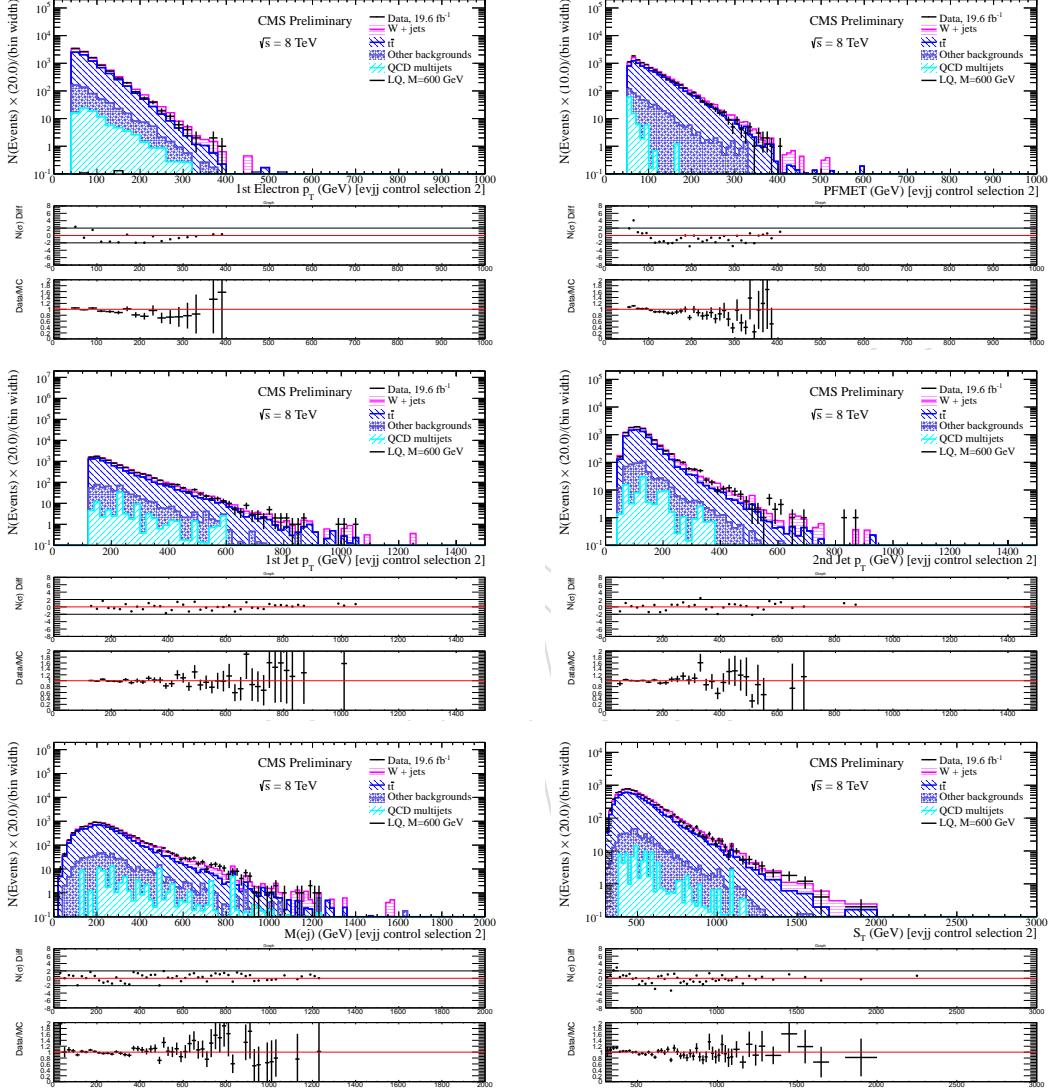


Figure 27: The electron  $p_T$  (top left),  $E_T^{\text{miss}}$  (top right), leading jet  $p_T$  (middle left), second leading jet  $p_T$  (middle right),  $m_{\text{ej}}$  (bottom left), and  $S_T$  (bottom right) distributions for an  $evjj$  sample enriched in  $t\bar{t}$  events (selection 2) after the rescaling of this MC background simultaneously with the rescaling of the  $W+jets$  background.

## 582 6 Systematic uncertainties

### 583 6.1 Background normalization

584 The uncertainties on the normalization / rescaling factors of the main backgrounds  
 585 are discussed in Section 5 and summarized below:

- 586 • overall uncertainty on QCD multijet background in the  $eejj$  ( $evjj$ ) channel:  
 587    60% (30%);
- 588 • scale factor for  $e\mu jj$  sample for  $t\bar{t}$  background estimate in the  $eejj$  channel:  
 589    Table 18;
- 590 • scale factor for  $t\bar{t}$  MC sample in the  $evjj$  channel:  
 591     $\mathcal{R}_{t\bar{t}} = 0.97 \pm 0.02$  (stat)  $\pm 0.01$  (syst);
- 592 • scale factor for  $Z^0$ +jets MC sample in the  $eejj$  channel:  
 593     $\mathcal{R}_{Z^0 + \text{jets}} = 0.97 \pm 0.01$  (stat)  $\pm 0.00004$  (syst);
- 594 • scale factor for  $W$ +jets MC sample in the  $evjj$  channel:  
 595     $\mathcal{R}_{W + \text{jets}} = 0.85 \pm 0.01$  (stat)  $\pm 0.01$  (syst);

### 596 6.2 $t\bar{t}$ , $Z^0$ +jets, $W$ +jets background shape

597 The systematic uncertainties due to the modeling of the  $Z^0$ +jets background (in the  
 598  $eejj$  channel), and both  $t\bar{t}$  and  $W$ +jets backgrounds (in the  $evjj$  channel) are deter-  
 599 mined by comparing the background predictions obtained with the default MC sam-  
 600 ples listed in Section 2.2 with those derived using alternative MADGRAPH samples  
 601 with normalization/factorization scales and jet matching thresholds varied by a fac-  
 602 tor of two. These alternative MC samples are summarized in Table 19.

Table 19: Alternative Monte Carlo datasets used to model SM backgrounds, with normalization/factorization scales and jet matching thresholds varied by a factor of two

Dataset	Process
/TTJets_matchingdown_TuneZ2star_8TeV-madgraph-tauola/ /TTJets_matchingup_TuneZ2star_8TeV-madgraph-tauola/ /TTJets_scaledown_TuneZ2star_8TeV-madgraph-tauola/ /TTJets_scaleup_TuneZ2star_8TeV-madgraph-tauola/	$t\bar{t}$
/WJetsToLNu_matchingdown_8TeV-madgraph-tauola/ /WJetsToLNu_matchingup_8TeV-madgraph-tauola/ /WJetsToLNu_scaledown_8TeV-madgraph-tauola/ /WJetsToLNu_scaleup_8TeV-madgraph-tauola/	$W + \text{jets}$
/DYJetsToLL_M-50_matchingdown_8TeV-madgraph/ /DYJetsToLL_M-50_matchingup_8TeV-madgraph-tauola/ /DYJetsToLL_M-50_scaledown_8TeV-madgraph-tauola/ /DYJetsToLL_M-50_scaleup_8TeV-madgraph-tauola/	$Z^0 + \text{jets}$

603 The method for evaluating these uncertainties has been taken from the second gen-  
 604 eration leptoquark analysis, and it is summarized below. All samples (the nominal  
 605 and alternative samples) are normalized at the preselection value. The lowest final  
 606 selection cuts are then applied. The alternative sample for a given physics process  
 607 with the maximal percentage difference in acceptance from the central sample is  
 608 identified. This percentage difference is taken as the shape systematic for all final  
 609 selection points. Because statistics in these alternative samples are limited, it is not  
 610 possible to repeat this method at tighter final selection levels. Furthermore, for the

611 W and Z alternative samples, the  $m_{ee}$  and  $m_{T, ev}$  cuts are removed from the final  
 612 selection.

613 The shape systematics determined with the above method are:

- 614 •  $Z^0 + \text{jets}$  in the  $eejj$  channel: 8.39%
- 615 •  $W + \text{jets}$  in the  $evjj$  channel: 7.50%
- 616 •  $t\bar{t}$  in the  $evjj$  channel: 5.64%

### 617 6.3 Electron, jet, and $E_T^{\text{miss}}$ energy scales

The electron energy scale uncertainty in the ECAL barrel (endcaps) is estimated to be 0.4% (4.1%) [31]. The jet energy scale uncertainties are evaluated for each jet using information from the global tags listed in Table 7. For the jet and electron energy scales, the event selection is repeated after rescaling the energy of the reconstructed objects by  $1 \pm \sigma$ , where  $\sigma$  is the relative uncertainty on their energy scales. While the electron and jet energy scale uncertainties are independent, the  $E_T^{\text{miss}}$  scale is directly correlated with both electron and jet energy scales. Therefore, while varying the electron and jet energy scales, a new  $E_T^{\text{miss}}$  vector is computed for each event:

$$E_T^{\text{miss}}' = E_T^{\text{miss}} + \sum_{\text{electrons OR jets}} (\vec{p}_T - \vec{p}'_T) \quad (11)$$

618 where  $\vec{p}_T$  is the  $p_T$  vector of the original electron or jet,  $\vec{p}'_T$  is the  $p_T$  vector of the  
 619 electron or jet with modified energy scale, and the sum is performed over the leading  
 620 (in  $p_T$ ) 15 jets with  $|\eta| < 2.4$  and  $p_T > 10$  GeV or the leading (in  $p_T$ ) 10 electrons.

621 For the backgrounds with data-driven normalization, the same rescaling procedure  
 622 adopted for the default MC sample is repeated for the samples with varied elec-  
 623 tron or jet energy scales. In this way, we can evaluate the impact of the electron  
 624 and jet energy scale systematic uncertainties on the final selection, but only relative  
 625 to the preselection stage. The maximum percentage change in the total number of  
 626 predicted signal and background events, at each stage of final selection, due to vari-  
 627 ations of the jet or electron energy scales is used to assess the (symmetric) systematic  
 628 uncertainties.

629 These uncertainties are not considered for the QCD multijet background in both  
 630  $eejj$  and  $evjj$  channels, since that background is derived from data.

### 631 6.4 Electron and jet energy resolution

The effect of the jet energy resolution uncertainty in MC is calculated by matching  
 GEN-jets to RECO-jets and then reassigning the  $p_T$  of the RECO object, according to  
 Equation 12:

$$p'_{T, \text{RECO}} = p_{T, \text{GEN}} + c \cdot (p_{T, \text{RECO}} - p_{T, \text{GEN}}) \quad (12)$$

632 where  $c$  is the ratio of the jet energy resolution in data vs. MC. This ratio has been  
 633 found to depend on pseudorapidity, and the values for various regions are shown  
 634 in Table 20.

635 For the backgrounds with data-driven normalization, the same rescaling procedure  
 636 adopted for the default MC sample is repeated for the samples with varied jet energy  
 637 resolution. In this way, we can evaluate the impact of the jet energy resolution sys-  
 638 tematic uncertainty on the final selection, but only relative to the preselection stage.  
 639 The maximum percentage change in the total number of predicted signal and back-  
 640 ground events, at each stage of final selection, due to variation of the jet resolution  
 641 is used to assess the (symmetric) systematic uncertainties.

Table 20: Ratio between data and MC values for jet energy resolution for different pseudorapidity regions of the detector, as measured by the JetMET group

$ \eta $ range	Data / MC ratio
0.0 - 0.5	$1.052 \pm 0.012$ (stat) $\pm 0.062$ (syst)
0.5 - 1.1	$1.057 \pm 0.012$ (stat) $\pm 0.056$ (syst)
1.1 - 1.7	$1.096 \pm 0.017$ (stat) $\pm 0.063$ (syst)
1.7 - 2.3	$1.134 \pm 0.035$ (stat) $\pm 0.087$ (syst)
2.3 - 5.0	$1.288 \pm 0.127$ (stat) $\pm 0.155$ (syst)

642 The effect of the electron energy resolution uncertainty in MC is calculated using the  
 643 same method as for jets, but the value for  $c$  is taken to be 1.006 in the barrel and  
 644 1.015 in the endcap [21]. All of the changes to the electron  $p_T$  are propagated to the  
 645  $E_T^{\text{miss}}$ .

646 These uncertainties are not considered for the QCD multijet background in both  
 647  $eejj$  and  $evjj$  channels, since that background is derived from data.

## 6.5 Integrated luminosity

649 The uncertainty on the luminosity is 2.6%.

## 6.6 MC statistics

651 The statistical uncertainty on the number of MC events is summarized after the full  
 652  $eejj$  ( $evjj$ ) event selection in Table 11 (Table 13).

## 6.7 Electron trigger, reconstruction, identification, and isolation uncertainties

655 Electron efficiency may be factorized into three components:  $\epsilon_{\text{ele}} = \epsilon_{\text{gsf}} \cdot \epsilon_{\text{HEEP}} \cdot$   
 656  $\epsilon_{\text{trigger}}$ . The values and uncertainties for these components are summarized in Table  
 657 21.

658  $\epsilon_{\text{gsf}}$  refers to the fraction of electrons forming an ECAL supercluster (energy deposit  
 659 in the ECAL) that are reconstructed as gsf electrons.

660  $\epsilon_{\text{HEEP}}$  refers to the HEEP v4.1 electron ID efficiency, which has been measured by  
 661 the HEEP group using a tag and probe study. Tag electrons are required to be  
 662 barrel electrons ( $|\eta| < 1.442$ ) passing the HEEP v4.1 electron ID. Probe electrons  
 663 are required to be gsf electrons, and the combined invariant mass with a tag elec-  
 664 tron must be in the range  $60 < m_{\text{ee}} < 120$  GeV. Passing probe electrons are re-  
 665 quired to pass the HEEP v4.1 electron ID. Events for the study are selected using the  
 666 `HLT_DoubleEle33_CaloIdL_GsfTrkIdVL` trigger.

667  $\epsilon_{\text{trigger}}$  refers to the efficiency with which electrons that pass the HEEP v4.1 electron  
 668 ID also fire the `Ele30_CaloIdVT_TrkIdT` leg of the trigger (see Tables 8 and 9).  
 669 Since our offline electron  $p_T$  cut (45 GeV) is much higher than the online electron  $p_T$   
 670 cut required by the trigger (30 GeV), we assume that our selection is within the effi-  
 671 ciency plateau region for this trigger, and we use the plateau efficiency calculated by  
 672 the  $W'$  analysis [30] with which HEEP v4.1 electrons also fire the `HLT_Ele80_CaloIdVT_TrkIdT`  
 673 trigger. This efficiency has been calculated for data only. Our assumption is sup-  
 674 ported by the good agreement between data (on which we apply the trigger) and Monte Carlo  
 675 (on which we do not apply the trigger) at low electron  $p_T$  in Figures 4 and 14.

676 Because the gsf reconstruction efficiency scale factor and the HEEP v4.1 electron ID  
 677 efficiency scale factor agree with unity, we do not apply any scale factor to the Monte

Table 21: Efficiencies on electron reconstruction, the HEEP v4.1 electron ID, and the electron leg of the signal triggers in Tables 8 and 9 in both data and Monte Carlo. The gsf efficiencies are provided by the egamma POG. The HEEP efficiencies are taken from [32]. The uncertainties on the HEEP scale factors are taken from [33]. The electron trigger efficiencies are taken from the  $W'$  analysis [30].

Efficiency	Data	Simulation	Data / Simulation
Barrel			
gsf reco	98.5%	98.9%	$0.995 \pm 0.012$
HEEP v4.1 [32]	$88.7\% \pm 0.2\%$ (stat.)	$90.2\% \pm 0.2\%$ (stat.)	$0.98 \pm 0.02$
Trigger Ele30	$97.4\% \pm 0.56\%$	-	-
Endcap			
gsf reco	95.0%	95.6%	$0.994 \pm 0.015$
HEEP v4.1 [32]	$90.7\% \pm 0.6\%$ (stat.)	$92.2\% \pm 0.5\%$ (stat.)	$0.98 \pm 0.04$
Trigger Ele30	$95.8\% \pm 1.35\%$	-	-

678 Carlo to account for these effects. We take the deviation from unity as a systematic  
 679 uncertainty. The per-electron trigger efficiency,  $\epsilon_{\text{trigger}}$ , which does not agree with  
 680 unity, is applied to the signal Monte Carlo.

## 681 6.8 Parton distribution function (PDF)

682 **FIXME: not yet implemented in the analysis**

## 683 6.9 Pileup

684 The uncertainty due to the modeling of the pileup interactions is estimated by re-  
 685 weighting the MC events to match with a number of pileup events 6% larger or  
 686 smaller than what is observed in data, accordingly with the uncertainty on the total  
 687 inelastic pp cross section. The number of events passing the full selection is calcu-  
 688 lated for the  $\pm 6\%$  cases and the default one. The maximum variation with respect  
 689 to the default is used to asses a systematic uncertainty. The uncertainty is less than  
 690 1% for background and LQ signal at all masses in both the  $eejj$  and  $evjj$  channels.

Table 22: Systematic uncertainties and their effects on signal and background in the  $eejj$  channel for  $M(\text{LQ}) = 300$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	4.00%	0.00%
Electron energy scale	0.57%	0.80%
Electron energy resolution	0.06%	0.02%
Jet energy scale	1.10%	0.34%
Jet energy resolution	0.05%	0.16%
Pileup	0.14%	0.23%
Luminosity	2.60%	0.09%
Z normalization	0.00%	0.41%
Z shape	0.00%	3.21%
t̄t estimate	0.00%	1.16%
QCD multijet estimate	0.00%	0.13%
Total	4.93%	3.56%

Table 23: Systematic uncertainties and their effects on signal and background in the  $eejj$  channel for  $M(LQ) = 500$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	4.00%	0.00%
Electron energy scale	0.41%	1.18%
Electron energy resolution	0.00%	0.03%
Jet energy scale	0.42%	0.48%
Jet energy resolution	0.01%	0.22%
Pileup	0.18%	0.33%
Luminosity	2.60%	0.12%
Z normalization	0.00%	0.60%
Z shape	0.00%	4.76%
$t\bar{t}$ estimate	0.00%	0.78%
QCD multijet estimate	0.00%	0.12%
Total	4.81%	5.04%

Table 24: Systematic uncertainties and their effects on signal and background in the  $eejj$  channel for  $M(LQ) = 700$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	4.00%	0.00%
Electron energy scale	0.41%	1.35%
Electron energy resolution	0.02%	0.04%
Jet energy scale	0.36%	0.43%
Jet energy resolution	0.00%	0.19%
Pileup	0.07%	0.34%
Luminosity	2.60%	0.04%
Z normalization	0.00%	0.70%
Z shape	0.00%	5.57%
$t\bar{t}$ estimate	0.00%	0.68%
QCD multijet estimate	0.00%	0.11%
Total	4.80%	5.85%

Table 25: Systematic uncertainties and their effects on signal and background in the  $eejj$  channel for  $M(LQ) = 900$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	4.00%	0.00%
Electron energy scale	0.39%	1.46%
Electron energy resolution	0.05%	0.04%
Jet energy scale	0.41%	0.47%
Jet energy resolution	0.02%	0.20%
Pileup	0.10%	0.36%
Luminosity	2.60%	0.03%
Z normalization	0.00%	0.76%
Z shape	0.00%	6.02%
$t\bar{t}$ estimate	0.00%	0.58%
QCD multijet estimate	0.00%	0.12%
Total	4.81%	6.30%

Table 26: Systematic uncertainties and their effects on signal and background in the  $eejj$  channel for  $M(LQ) = 1100$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	4.00%	0.00%
Electron energy scale	0.28%	2.08%
Electron energy resolution	0.01%	0.06%
Jet energy scale	0.31%	0.66%
Jet energy resolution	0.00%	0.29%
Pileup	0.27%	0.52%
Luminosity	2.60%	0.01%
Z normalization	0.00%	1.09%
Z shape	0.00%	8.60%
$t\bar{t}$ estimate	0.00%	0.00%
QCD multijet estimate	0.00%	0.20%
Total	4.80%	8.97%

Table 27: Systematic uncertainties and their effects on signal and background in the  $e\nu jj$  channel for  $M(LQ) = 300$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	2.00%	0.00%
Electron energy scale	0.72%	2.02%
Electron energy resolution	0.24%	0.87%
Jet energy scale	2.19%	3.33%
Jet energy resolution	0.24%	0.76%
Pileup	0.31%	1.68%
Luminosity	2.60%	0.15%
W normalization	0.00%	0.75%
W shape	0.00%	2.87%
$t\bar{t}$ normalization	0.00%	1.51%
$t\bar{t}$ shape	0.00%	3.01%
QCD multijet estimate	0.00%	1.51%
Total	4.04%	6.46%

Table 28: Systematic uncertainties and their effects on signal and background in the  $e\nu jj$  channel for  $M(LQ) = 500$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	2.00%	0.00%
Electron energy scale	0.96%	1.65%
Electron energy resolution	0.04%	0.73%
Jet energy scale	1.31%	2.71%
Jet energy resolution	0.21%	0.61%
Pileup	0.23%	1.38%
Luminosity	2.60%	0.33%
W normalization	0.00%	0.56%
W shape	0.00%	2.15%
$t\bar{t}$ normalization	0.00%	1.31%
$t\bar{t}$ shape	0.00%	2.61%
QCD multijet estimate	0.00%	2.98%
Total	3.67%	5.95%

Table 29: Systematic uncertainties and their effects on signal and background in the  $e\nu jj$  channel for  $M(LQ) = 700$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	2.00%	0.00%
Electron energy scale	0.78%	1.52%
Electron energy resolution	0.06%	0.74%
Jet energy scale	1.40%	2.34%
Jet energy resolution	0.12%	0.53%
Pileup	0.29%	1.29%
Luminosity	2.60%	0.26%
W normalization	0.00%	0.20%
W shape	0.00%	0.75%
$t\bar{t}$ normalization	0.00%	1.74%
$t\bar{t}$ shape	0.00%	3.47%
QCD multijet estimate	0.00%	5.74%
Total	3.67%	7.68%

Table 30: Systematic uncertainties and their effects on signal and background in the  $e\nu jj$  channel for  $M(LQ) = 900$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	2.00%	0.00%
Electron energy scale	0.81%	1.76%
Electron energy resolution	0.04%	0.85%
Jet energy scale	1.89%	2.68%
Jet energy resolution	0.07%	0.59%
Pileup	0.53%	1.48%
Luminosity	2.60%	0.40%
W normalization	0.00%	0.00%
W shape	0.00%	0.00%
$t\bar{t}$ normalization	0.00%	2.07%
$t\bar{t}$ shape	0.00%	4.14%
QCD multijet estimate	0.00%	4.28%
Total	3.91%	7.31%

Table 31: Systematic uncertainties and their effects on signal and background in the  $evjj$  channel for  $M(LQ) = 1100$  GeV final selection. All uncertainties are symmetric.

Systematic	Signal (%)	Background (%)
Electron efficiency	2.00%	0.00%
Electron energy scale	0.59%	1.78%
Electron energy resolution	0.15%	0.86%
Jet energy scale	1.07%	2.71%
Jet energy resolution	0.02%	0.60%
Pileup	0.12%	1.50%
Luminosity	2.60%	0.40%
W normalization	0.00%	0.00%
W shape	0.00%	0.00%
t̄t normalization	0.00%	2.09%
t̄t shape	0.00%	4.19%
QCD multijet estimate	0.00%	3.95%
Total	3.50%	7.18%

## 691 7 Results

692 As shown in Table 11 and Table 13, both channels exhibit a broad excess with re-  
 693 spect to the predicted contribution from Standard Model background in nearly all  
 694 of the levels of final selection, despite good agreement between observed events and  
 695 the Standard Model background prediction at preselection. The  $p$ -values and sig-  
 696 nificances for these discrepancies with respect to the Standard Model background  
 697 prediction as a function of the hypothesized leptoquark mass are calculated using  
 698 asymptotic CLs and shown in Figure 28. The significances are evaluated before any  
 699 “look elsewhere” effect is considered. From Figure 28, it is clear that the discrepancy  
 700 is most significant for both analyses in the final selection optimized for a leptoquark  
 701 of mass 650 GeV.

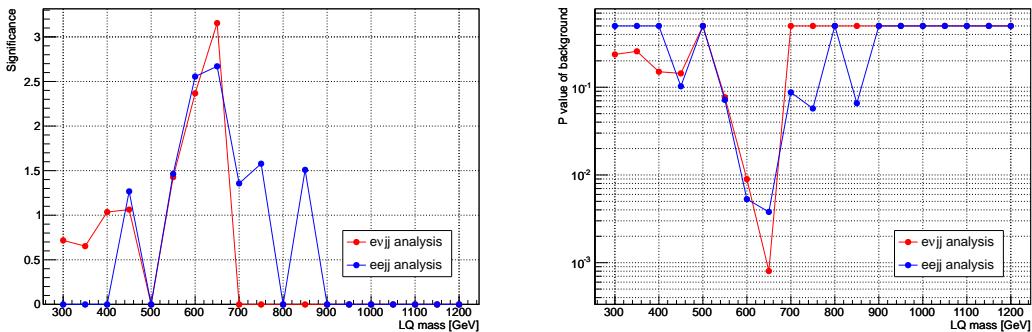


Figure 28: The excess significance (left plot) and p value of the background (right plot) as a function of leptoquark mass.

702 However, Figures 11 and 12 make clear that the excess events in the  $eejj$  channel do  
 703 not peak sharply in  $m_{ej}^{\min}$ , as would be expected for a leptoquark-like signal. Simi-  
 704 larly, Figures 22 and 23 make clear that the excess events in the  $evjj$  channel do not  
 705 peak sharply in  $m_{ej}$ .

706 The discrepancies in both channels are significantly less than what would be ex-  
 707 pected in the case of a leptoquark with a mass of 650 GeV and  $\beta \geq 0.5$ . For this

708 reason, it is still possible to set competitive limits on first generation leptoquarks for  
 709  $\beta = 1.0$  and  $\beta = 0.5$ . A log-normal probability function is used to integrate over  
 710 the systematic uncertainties described in Section 6. Statistical uncertainties are de-  
 711 scribed with  $\Gamma$  distributions, the widths of which are determined by the number of  
 712 events in simulation samples before any reweighting or by the number of events ob-  
 713 served in data control regions. This method is also used by the second generation  
 714 leptoquark analysis.

715 Figure 29 shows the 95% CL upper limits on  $\sigma \times \beta^2$  ( $\sigma \times 2\beta(1 - \beta)$ ) as a function of  
 716 leptoquark mass for scalar leptoquarks in the  $eejj$  ( $evjj$ ) channel. The solid black line  
 717 denotes the observed limit, the dashed black line denotes the central value of the  
 718 predicted limit, and the green and yellow bands denote the systematic uncertainty  
 719 within one and two standard deviations of the central value of the predicted limit,  
 720 respectively. The theoretical cross sections are shown as a blue line. A lighter blue  
 721 band around that line denotes the combined theoretical uncertainty coming from  
 722 both the factorization and PDF uncertainties, as shown in Figure 5. The excess in  
 723 both analyses is clearly visible for leptoquarks with a mass of 650 GeV. Using Fig-  
 724 ure 29, 95% CL exclusion limits may be placed on first generation scalar leptoquarks  
 725 with masses less than 1005(845) GeV, assuming  $\beta = 1.0(0.5)$ . This is to be com-  
 726 pared with expected 95% CL exclusions of 1030(890) GeV, assuming  $\beta = 1.0(0.5)$ .  
 727 The limits in Figures 29 and 30 the p-values in Figure 28 were calculated using the  
 728 asymptotic CLs tool from the Higgs group<sup>3</sup>.

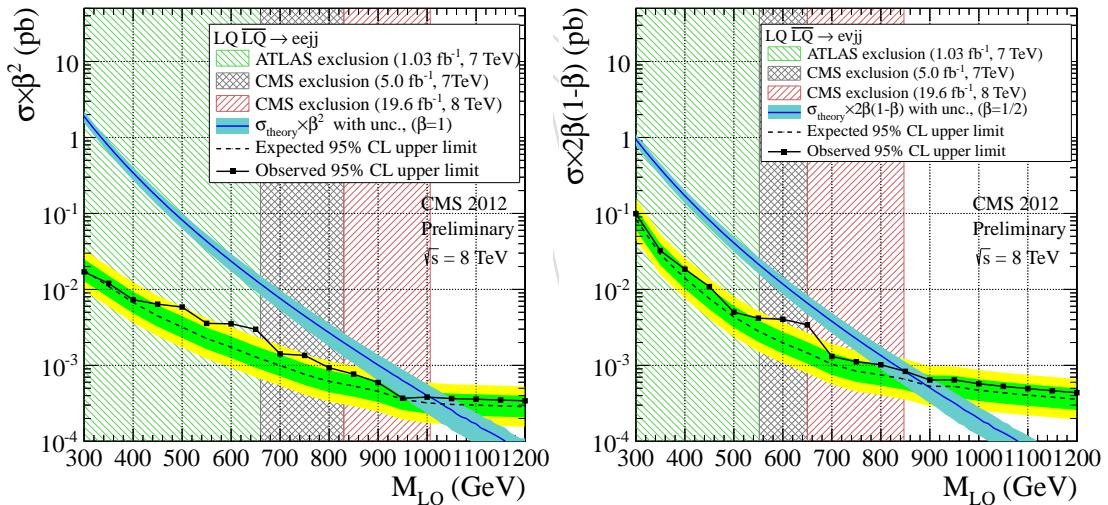


Figure 29: The expected and observed upper limit at 95% C.L. on the LQ pair production cross section times  $\beta^2$  in the top plot ( $2\beta(1 - \beta)$  in the bottom plot) as a function of the LQ mass obtained with the  $eejj$  ( $evjj$ ) analysis. The systematic uncertainties described in the text are included in the calculation. The dark blue curve and the light blue band represent, respectively, the theoretical LQ pair production cross section and the uncertainties due to the choice of PDF and renormalization/factorization scales.

729 The two independent limits from the  $eejj$  and  $evjj$  channels are combined in Figure  
 730 29. This combination results in a region of space in the  $\beta$ -mass plane which may  
 731 be excluded. This combination (denoted by a solid black line) is only affected sig-  
 732 nificantly by the excess in the  $evjj$  analysis (denoted by a solid red line). A leptoquark of

<sup>3</sup> Command used for p-values calculation: `combine -M ProfileLikelihood -n <label> -m <lq mass> --signif --pvalue -d <data_card>`

mass 650 GeV with  $\beta < 0.15$  cannot be excluded by this analysis. The combination does not change the observed limits for the case where  $\beta = 0.5$  or where  $\beta = 1.0$ . These limits for  $\beta = 0.5$  and  $\beta = 1.0$  represent the most stringent limits on first generation leptoquarks to date.

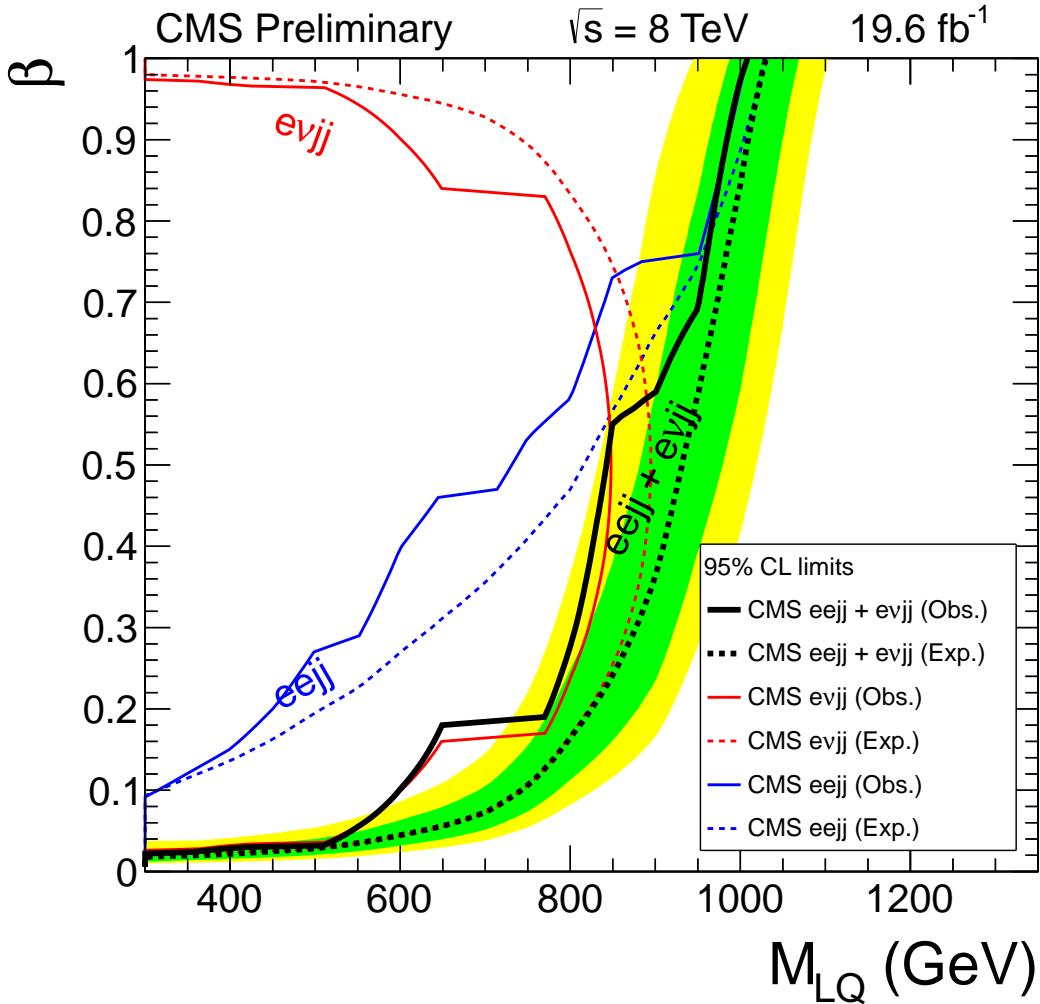


Figure 30: The expected and observed exclusion limits at 95% CL on the first generation leptoquark hypothesis in the  $\beta$  versus mass plane using the central value of signal cross section for the individual  $eejj$  and  $evjj$  channels and their combination. The green and yellow expected limit uncertainty bands represent the 68% and 95% confidence intervals. Solid lines represent the observed limits in each channel, and dashed lines represent the expected limits.

## References

- [1] H. Georgi and S. Glashow, "Unity of All Elementary-Particle Forces", *Phys. Rev. Lett.* **32** (1974) 438, doi:10.1103/PhysRevLett.32.438.
- [2] J. C. Pati and A. Salam, "Lepton Number as the Fourth Color", *Phys. Rev. D* **10** (1974) 275, doi:10.1103/PhysRevD.10.275.
- [3] B. Schrempp and F. Schrempp, "Light Leptoquarks", *Phys.Lett.* **B153** (1985) 101, doi:10.1016/0370-2693(85)91450-9.
- [4] S. Dimopoulos and L. Susskind, "Mass Without Scalars", *Nucl. Phys. B* **155** (1979) 237, doi:10.1016/0550-3213(81)90304-7.
- [5] S. Dimopoulos, "Technicolored Signatures", *Nucl. Phys. B* **168** (1980) 69, doi:10.1016/0550-3213(80)90277-1.
- [6] E. Eichten and K. Lane, "Dynamical Breaking of the Weak Interaction Symmetries", *Phys. Lett. B* **90** (1980) 85, doi:10.1016/0370-2693(80)90065-9.
- [7] J. L. Hewett and T. G. Rizzo, "Low-energy Phenomenology of Superstring-inspired  $E_6$  Models", *Physics Reports-review Section of Physics Letters* **183** (November, 1989) 193–381, doi:10.1016/0370-1573(89)90071-9.
- [8] CMS Collaboration Collaboration, "Search for pair production of first- and second-generation scalar leptoquarks in  $pp$  collisions at  $\sqrt{s} = 7$  TeV", *Phys. Rev. D* **86** (Sep, 2012) 052013, doi:10.1103/PhysRevD.86.052013.
- [9] M. Kuze and Y. Sirois, "Search for particles and forces beyond the standard model in high energy lepton-hadron and hadron-hadron collisions", *Progress in Particle and Nuclear Physics* **50** (2003), no. 1, 1 – 62, doi:10.1016/S0146-6410(02)00176-X.
- [10] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **86** (Jul, 2012) 010001, doi:10.1103/PhysRevD.86.010001. Leptoquarks are discussed on page 555.
- [11] W. Buchmüller and D. Wyler, "Constraints on SU(5)-type Leptoquarks", *Phys. Lett. B* **177** (1986) 377, doi:10.1016/0370-2693(86)90771-9.
- [12] GEANT4 Collaboration, "GEANT4: A simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [13] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.
- [14] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.
- [15] J. Alwall et al., "MadGraph/MadEvent v4: the new web generation", *JHEP* **09** (2007) 028, doi:10.1088/1126-6708/2007/09/028.
- [16] T. Sjstrand et al., "High-energy-physics event generation with Pythia6.1", *Computer Physics Communications* **135** (2001), no. 2, 238 – 259, doi:10.1016/S0010-4655(00)00236-8.
- [17] S. Alioli et al., "NLO vector-boson production matched with shower in POWHEG", *JHEP* **07** (2008) 060, doi:10.1088/1126-6708/2008/07/060, arXiv:0805.4802.

- [18] R. Gavin, Y. Li, F. Petriello, S. Quackenbush, “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order”, arXiv:1011.3540.
- [19] N. Kidonakis and R. Vogt, “Theoretical top quark cross section at the Fermilab Tevatron and the CERN LHC”, *Phys. Rev. D* **78** (Oct, 2008) 074005, doi:10.1103/PhysRevD.78.074005.
- [20] M. Krämer et al., “Pair production of scalar leptoquarks at the CERN LHC”, *Phys. Rev. D* **71** (Mar, 2005) 057503, doi:10.1103/PhysRevD.71.057503.
- [21] CMS Collaboration, “Search for Resonances in Dilepton Mass Spectra in pp Collisions at  $\sqrt{s} = 8$  TeV”, *CMS Physics Analysis Summary: EXO-12-061* (2013).
- [22] CMS Collaboration, “High Et electron selection presentation and optimisation in case of high pile-up conditions”, *CMS Physics Analysis Note: AN-2012/258* (2012).
- [23] “Performance of muon identification in pp collisions at  $\sqrt{s} = 7$  TeV”, Technical Report CMS-PAS-MUO-10-002, CERN, 2010. Geneva, (2010).
- [24] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k t jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063.
- [25] “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, (Apr, 2009).
- [26] “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, Technical Report CMS-PAS-PFT-10-001, 2010.
- [27] CMS Collaboration, “Particle Flow Jet Identification Criteria”, *CMS Physics Analysis Note: AN-2010/003* (2010).
- [28] CMS Collaboration, “Jet Identification”,.
- [29] “MET performance in 8 TeV data”, Technical Report CMS-PAS-JME-12-002, CERN, Geneva, (2013).
- [30] CMS Collaboration, “Search for new physics in the final states with a lepton and missing transverse energy at  $\sqrt{s} = 8$  TeV”, *CMS Physics Analysis Summary: EXO-12-060* (2013).
- [31] “ECAL Detector Performance, 2011 Data”,.
- [32] C. Clerbaux et al., “Search for High Mass Resonances Decaying to Electron Pairs at 8 TeV”, *CMS Note* **2012/415** (2012).
- [33] L. Thomas, “Update on the HEEP ID efficiency and scale factor”,.

816

## A $eejj$ analysis $M(LQ) = 650$ final selection extra plots

817

Distributions for the electron and jets in events passing the  $M(LQ) = 650$  final selection.

818

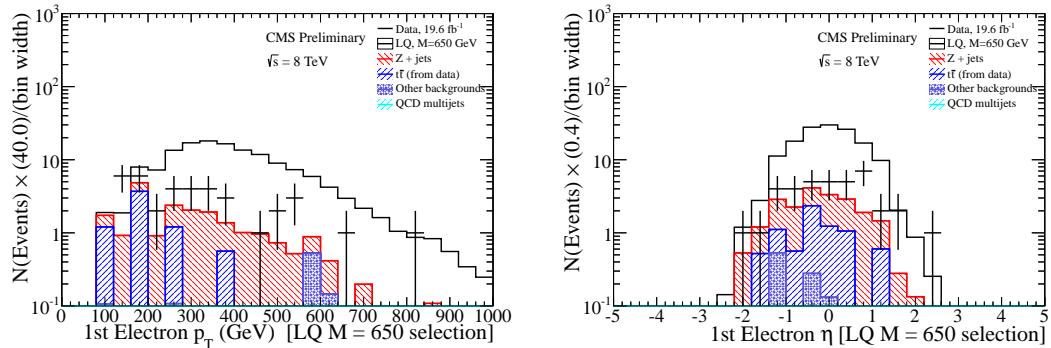


Figure 31: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $eejj$   $M(LQ) = 650$  final selection.

819

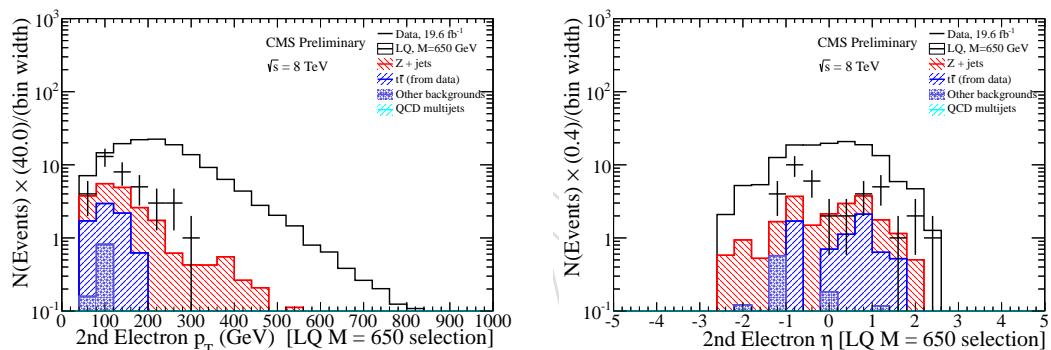


Figure 32: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) electron for events passing the  $eejj$   $M(LQ) = 650$  final selection.

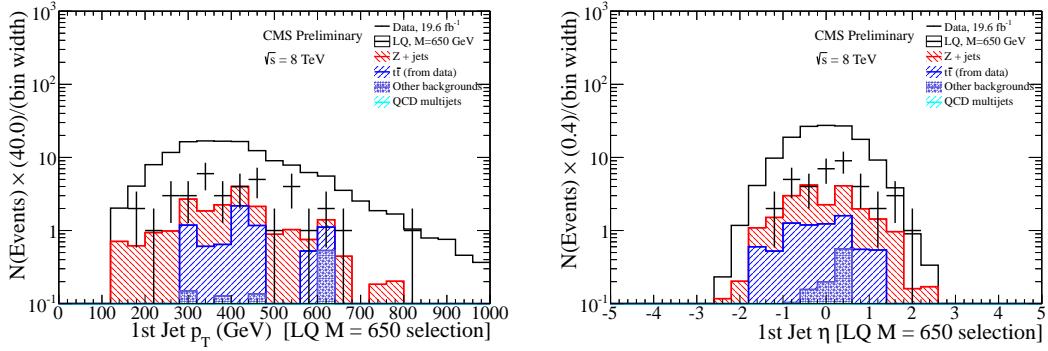


Figure 33: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $eejj$   $M(LQ) = 650$  final selection.

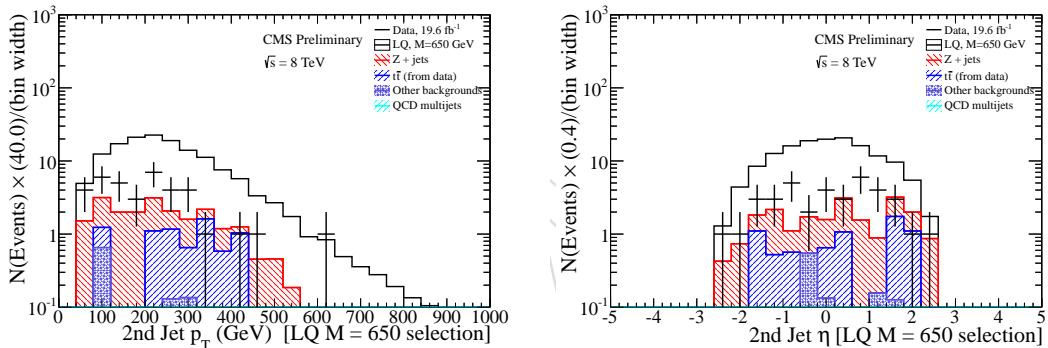


Figure 34: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $eejj$   $M(LQ) = 650$  final selection.

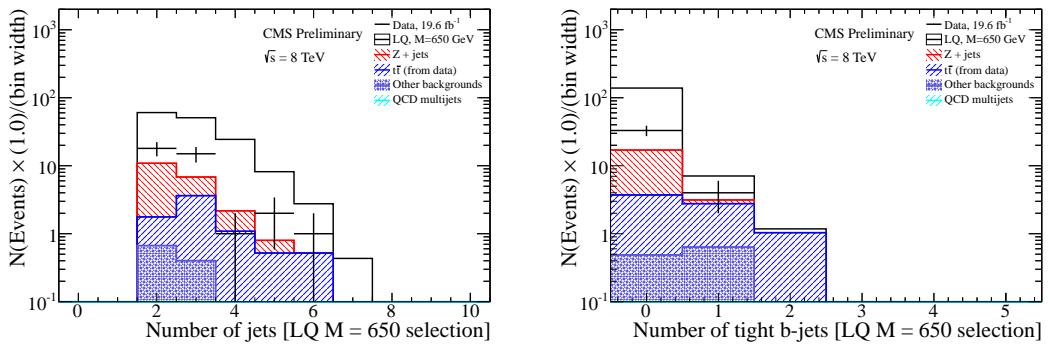


Figure 35: The number of jets with  $p_T > 45$  GeV (left) and the number of jets passing the tight CSV b-tag working point with  $p_T > 45$  GeV (right) for events passing the  $eejj$   $M(LQ) = 650$  final selection.

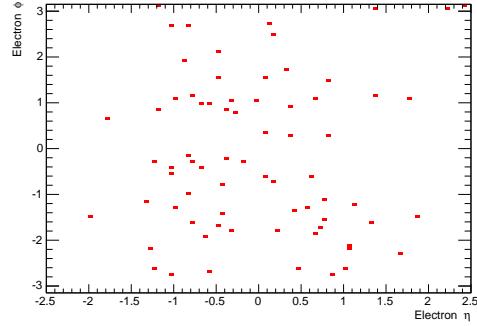


Figure 36: The  $\eta$  ( $x$ -axis) and  $\phi$  ( $y$ -axis) coordinates of all electrons passing the  $eejj$  M(LQ) = 650 final selection. There are two entries per event.

## B $eejj$ analysis M(LQ) = 650 final selection electron quality plots

Electron quality distributions for events passing the M(LQ) = 650 final selection of the  $eejj$  analysis.

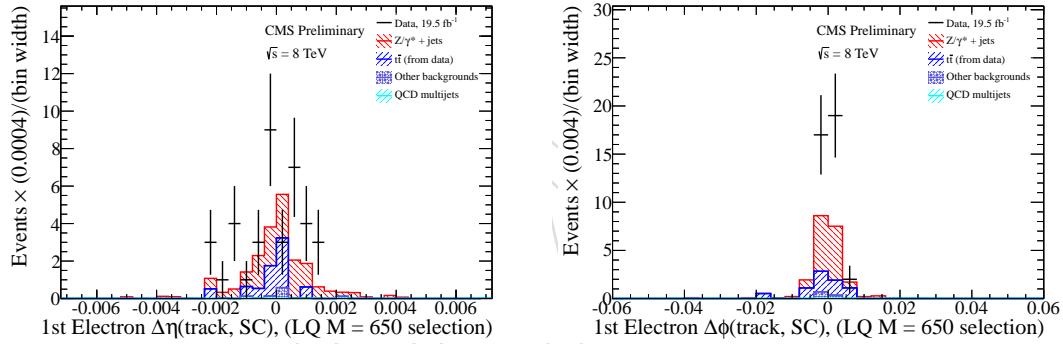


Figure 37: The  $\Delta\eta$  (left) and  $\Delta\phi$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $eejj$  M(LQ) = 650 final selection.

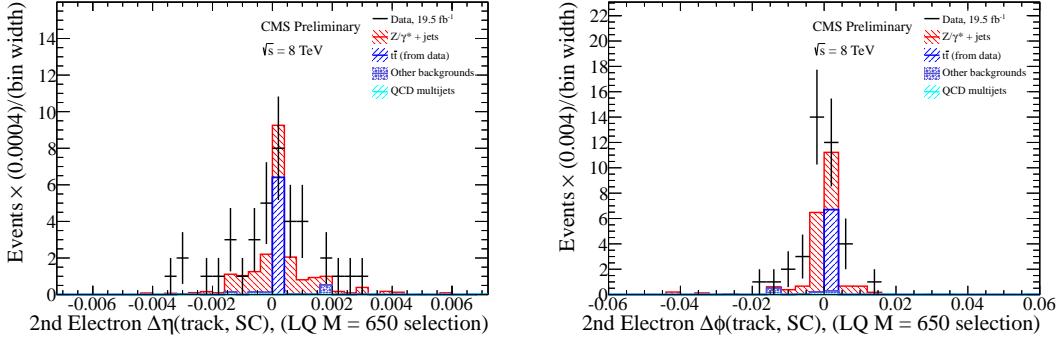


Figure 38: The  $\Delta\eta$  (left) and  $\Delta\phi$  (right) distributions of the second leading (in  $p_T$ ) electron for events passing the  $eejj$   $M(LQ) = 650$  final selection.

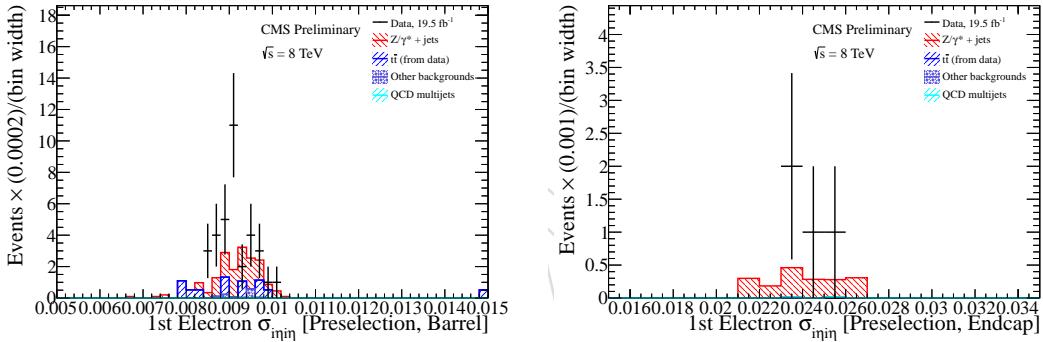


Figure 39: The  $\sigma_{i\eta i\eta}$  for the leading barrel electrons (left) and endcap electrons (right) for events passing the  $eejj$   $M(LQ) = 650$  final selection. This distribution is sensitive to spikes: low values of  $\sigma_{i\eta i\eta}$  (which do not appear in the above plots) could be an indication of hot towers.

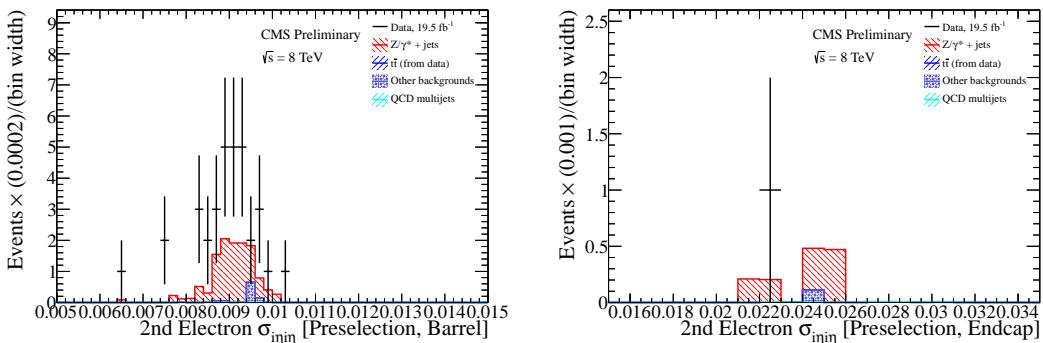


Figure 40: The  $\sigma_{i\eta i\eta}$  for the second leading barrel electrons (left) and endcap electrons (right) for events passing the  $eejj$   $M(LQ) = 650$  final selection. This distribution is sensitive to spikes: low values of  $\sigma_{i\eta i\eta}$  (which do not appear in the above plots) could be an indication of hot towers.

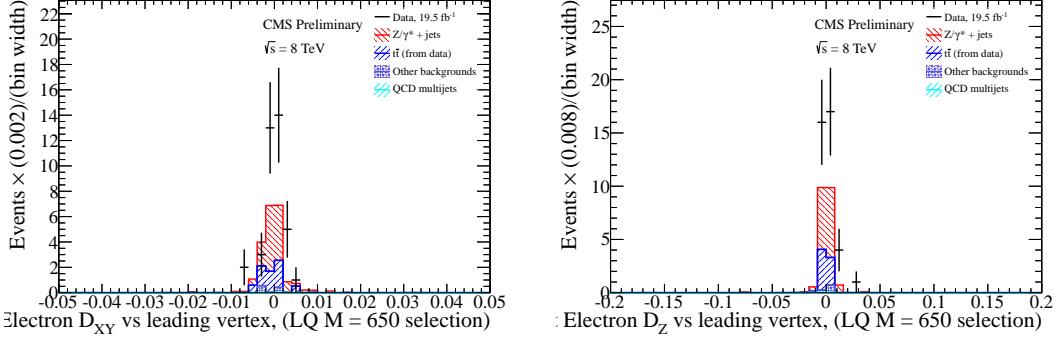


Figure 41: The transverse impact parameter (left) and longitudinal impact parameter (right) with respect to the leading primary vertex for leading electrons in events passing the  $eejj$   $M(LQ) = 650$  final selection. Shape discrepancies in these distributions could be a sign of contamination from QCD events or pileup collisions.

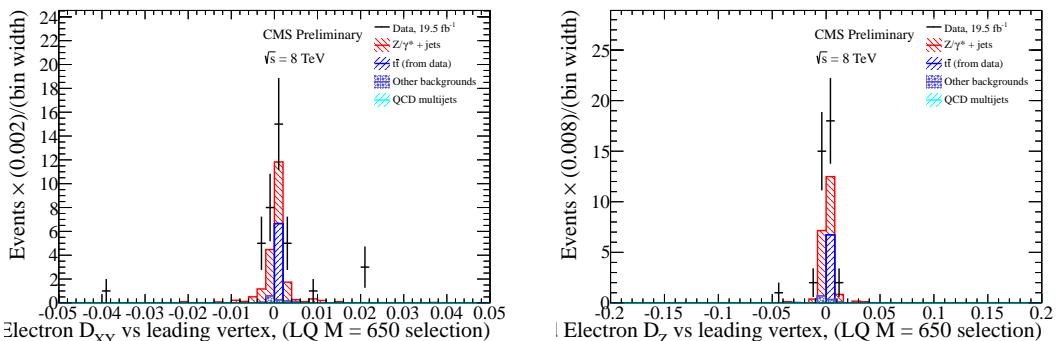


Figure 42: The transverse impact parameter (left) and longitudinal impact parameter (right) with respect to the leading primary vertex for second leading electrons in events passing the  $eejj$   $M(LQ) = 650$  final selection. Shape discrepancies in these distributions could be a sign of contamination from QCD events or pileup collisions.

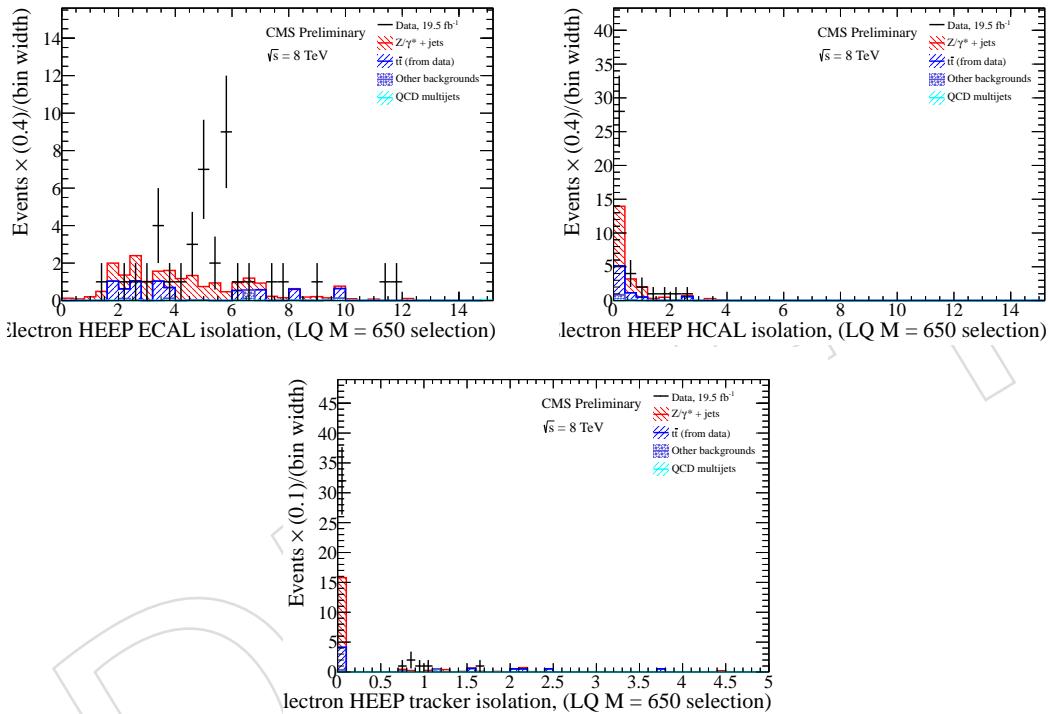


Figure 43: The ECAL isolation (top left), HCAL isolation (top right), and tracker isolation (bottom) for leading electrons in events passing the  $eejj$   $M(LQ) = 650$  final selection. Shape discrepancies in these distributions could be a sign of jet contamination in the electron isolation cone.

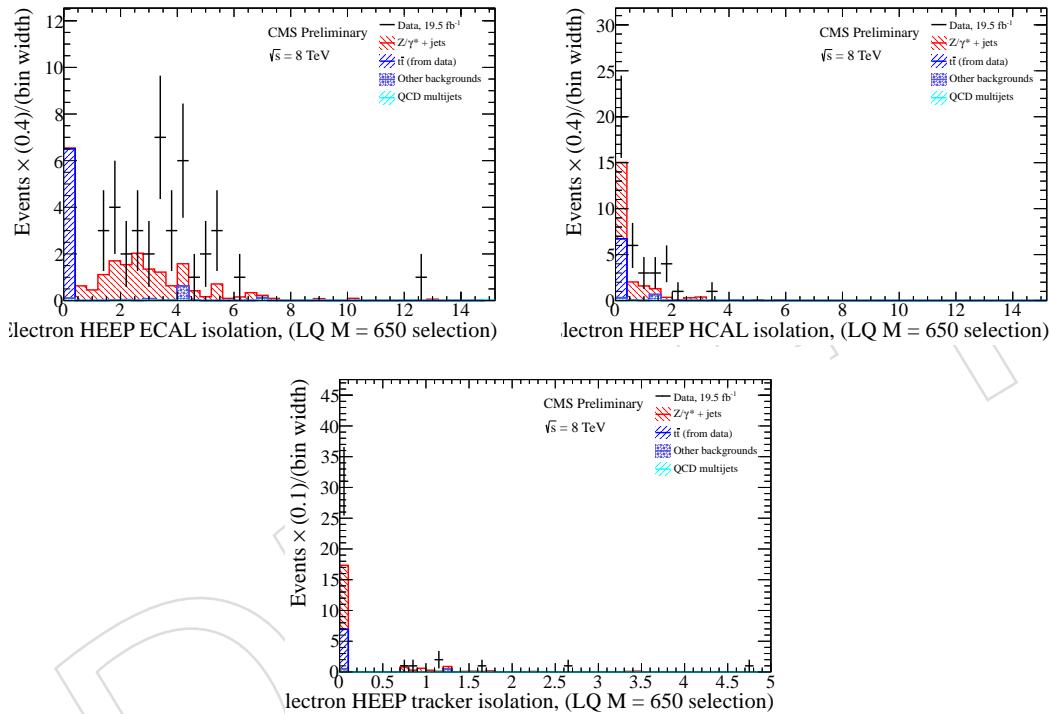


Figure 44: The ECAL isolation (top left), HCAL isolation (top right), and tracker isolation (bottom) for second leading electrons in events passing the  $eejj$   $M(\text{LQ}) = 650$  final selection. Shape discrepancies in these distributions could be a sign of jet contamination in the electron isolation cone.

824      **C  $eejj$  analysis  $M(LQ) = 650$  final selection invariant mass**  
 825      **plots**

826      Distributions for the invariant mass of various objects in events passing the  $M(LQ)$   
 827      = 650 final selection.

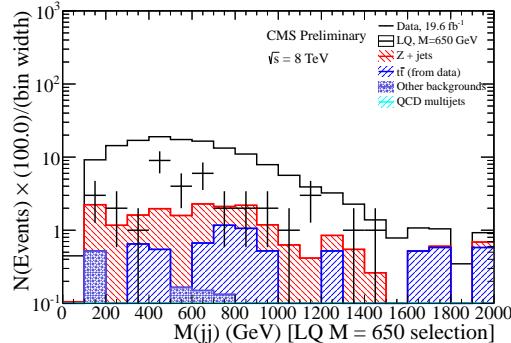


Figure 45: The invariant mass of the two leading jets in events passing the  $eejj$   $M(LQ) = 650$  final selection.

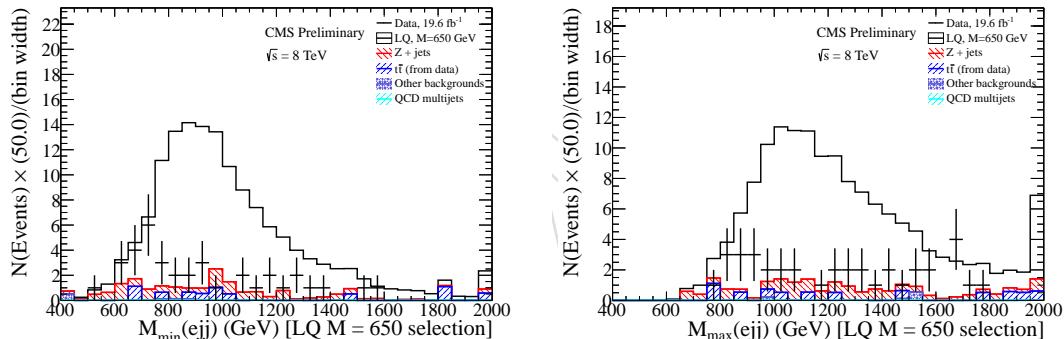


Figure 46: It is possible to make two electron + dijet combinations from the two leading electrons and two leading jets. The invariant mass distribution of the combination with the smaller invariant mass is shown on the left. The invariant mass distribution of the combination with the larger invariant mass is shown on the right.

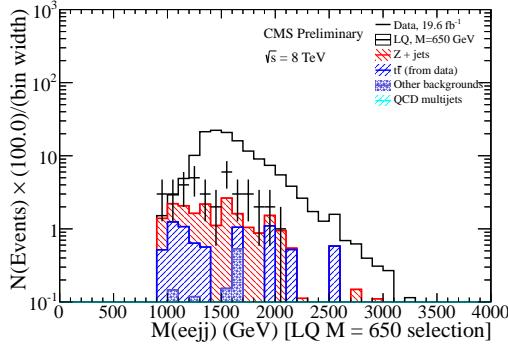


Figure 47: The invariant mass of the two leading electrons and two leading jets in events passing the  $eejj$   $M(LQ) = 650$  final selection.

## D $eejj$ analysis $M(LQ) = 650$ run period dependence

Distribution showing where in the 2012 run the events contributing to the  $M(LQ) = 450$  and  $650$  final selections were collected.

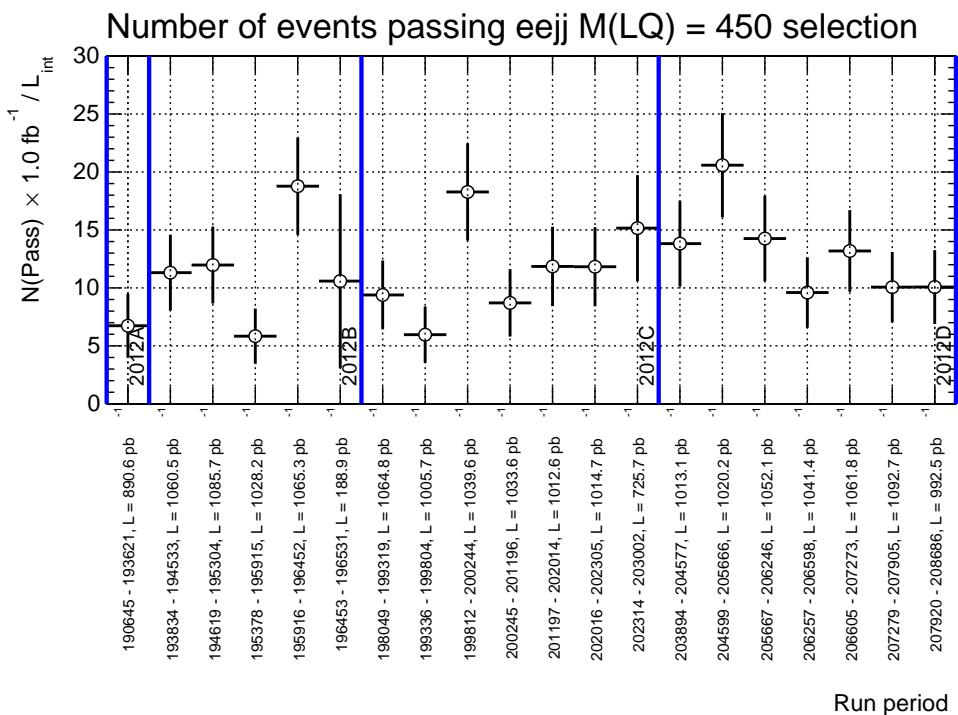


Figure 48: Distribution showing the number of events passing the  $M(LQ) = 450$  final selection of the  $eejj$  analysis during 20 run periods. Each run period corresponds to roughly  $1 \text{ fb}^{-1}$ . The minimum run, maximum run, and integrated luminosity of each run period is listed on the x-axis.

## E $evjj$ analysis $M(LQ) = 650$ final selection extra plots

Distributions for the electron and jets in events passing the  $M(LQ) = 650$  final selection.

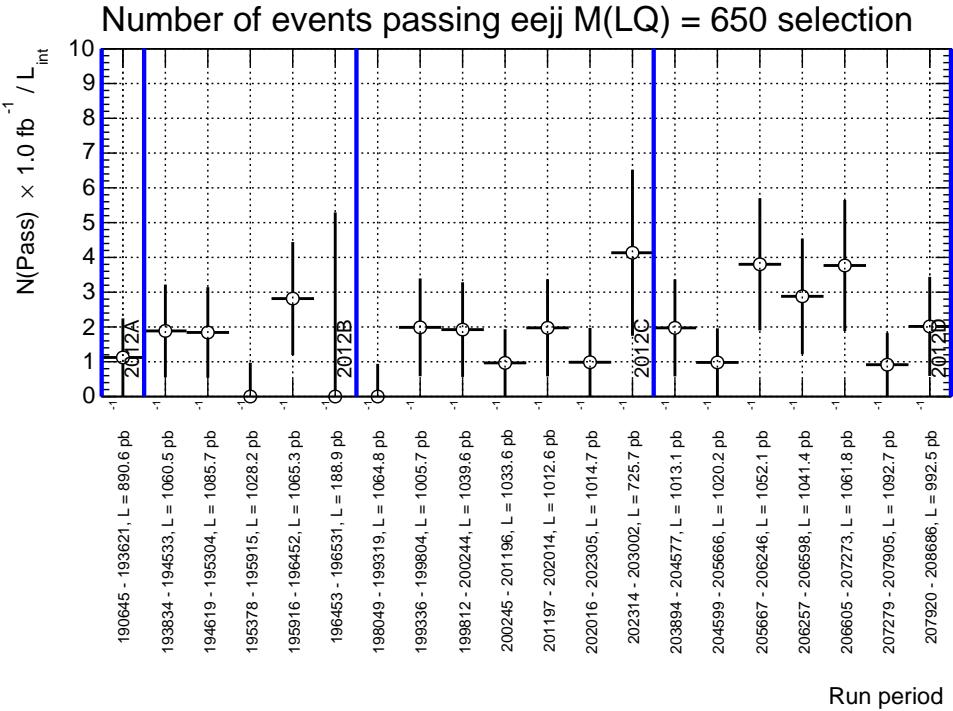


Figure 49: Distribution showing the number of events passing the  $M(\text{LQ}) = 650$  final selection of the  $eejj$  analysis during 20 run periods. Each run period corresponds to roughly  $1 \text{ fb}^{-1}$ . The minimum run, maximum run, and integrated luminosity of each run period is listed on the x-axis.

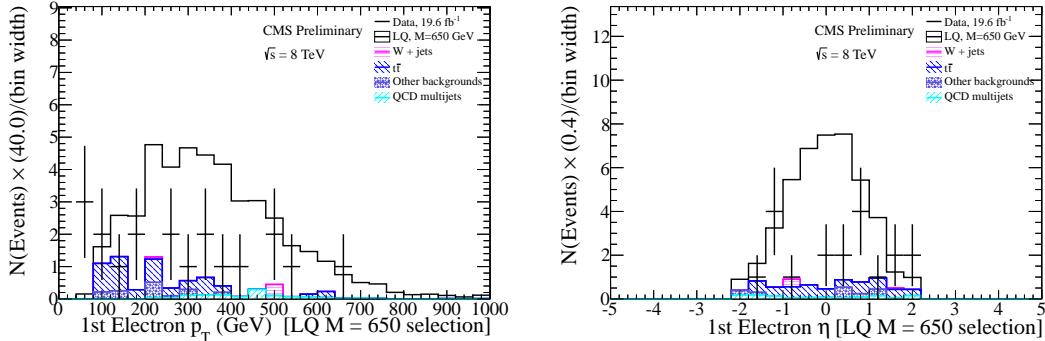


Figure 50: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $evjj$   $M(\text{LQ}) = 650$  final selection.

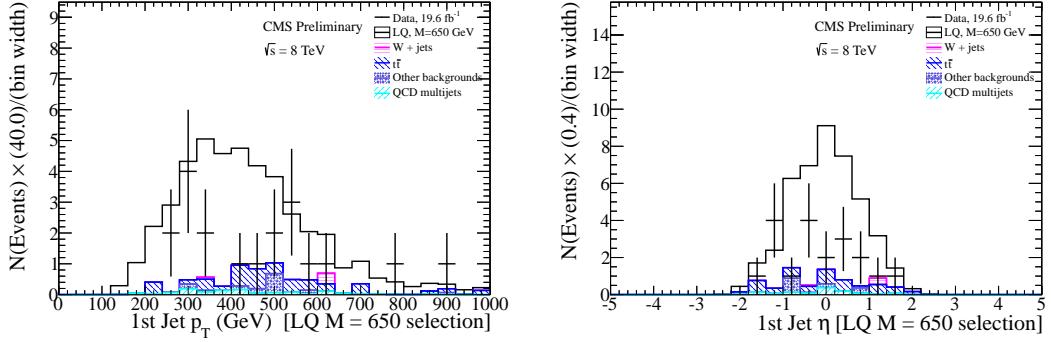


Figure 51: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $evjj$   $M(LQ) = 650$  final selection.

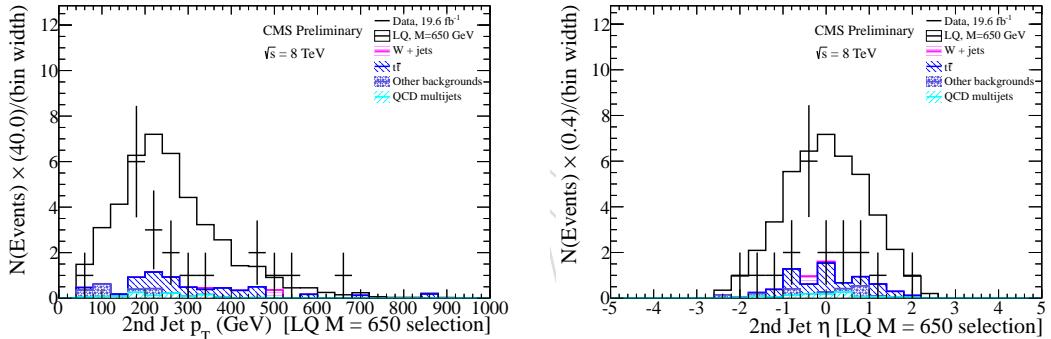


Figure 52: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $evjj$   $M(LQ) = 650$  final selection.

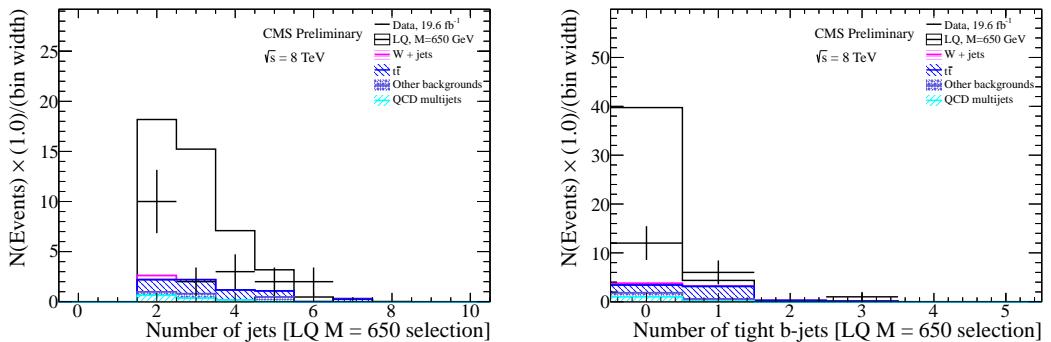


Figure 53: The number of jets with  $p_T > 45$  GeV (left) and the number of jets passing the tight CSV b-tag working point with  $p_T > 45$  GeV (right) for events passing the  $evjj$   $M(LQ) = 650$  final selection.

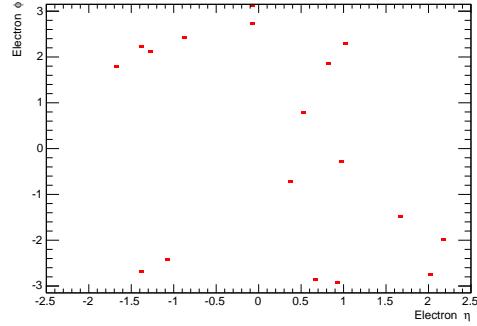


Figure 54: The  $\eta$  ( $x$ -axis) and  $\phi$  ( $y$ -axis) coordinates of all electrons passing the  $evjj$   $M(LQ) = 650$  final selection.

## 834      **F $evjj$ analysis $M(LQ) = 650$ final selection electron quality** 835      **plots**

836      Electron quality distributions for events passing the  $M(LQ) = 650$  final selection of  
 837      the  $evjj$  analysis.

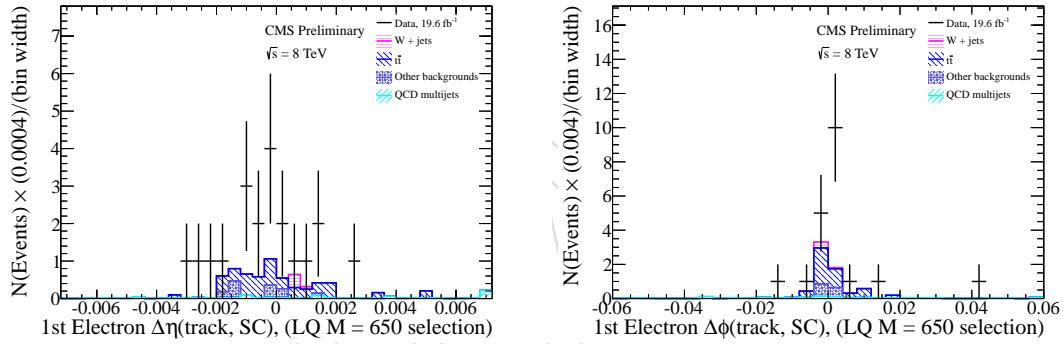


Figure 55: The  $\Delta\eta$  (left) and  $\Delta\phi$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $evjj$   $M(LQ) = 650$  final selection.

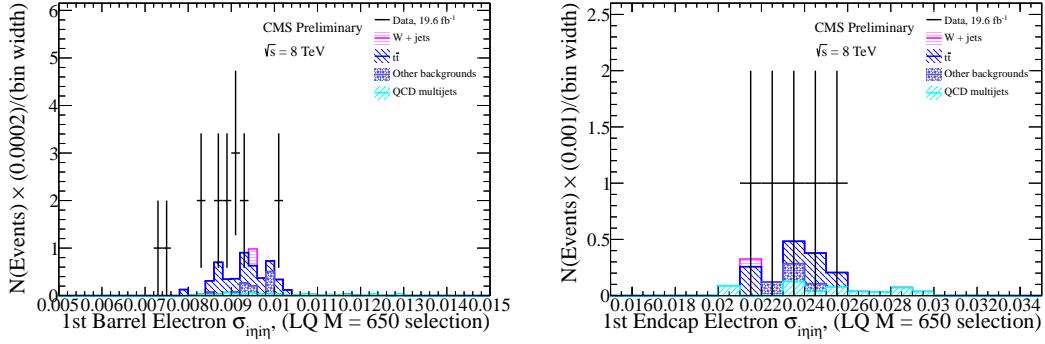


Figure 56: The  $\sigma_{i\eta i\eta}$  for barrel electrons (left) and endcap electrons (right) for events passing the  $evjj$   $M(LQ) = 650$  final selection. This distribution is sensitive to spikes: low values of  $\sigma_{i\eta i\eta}$  (which do not appear in the above plots) could be an indication of hot towers.

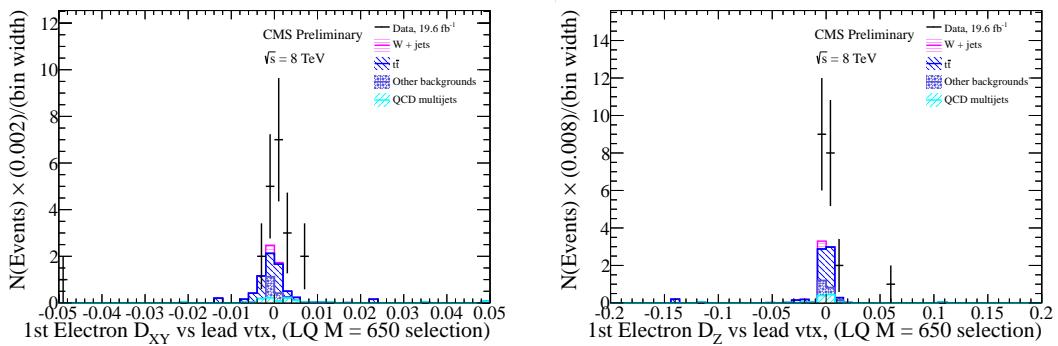


Figure 57: The transverse impact parameter (left) and longitudinal impact parameter (right) with respect to the leading primary vertex for electrons in events passing the  $evjj$   $M(LQ) = 650$  final selection. Shape discrepancies in these distributions could be a sign of contamination from QCD events or pileup collisions.

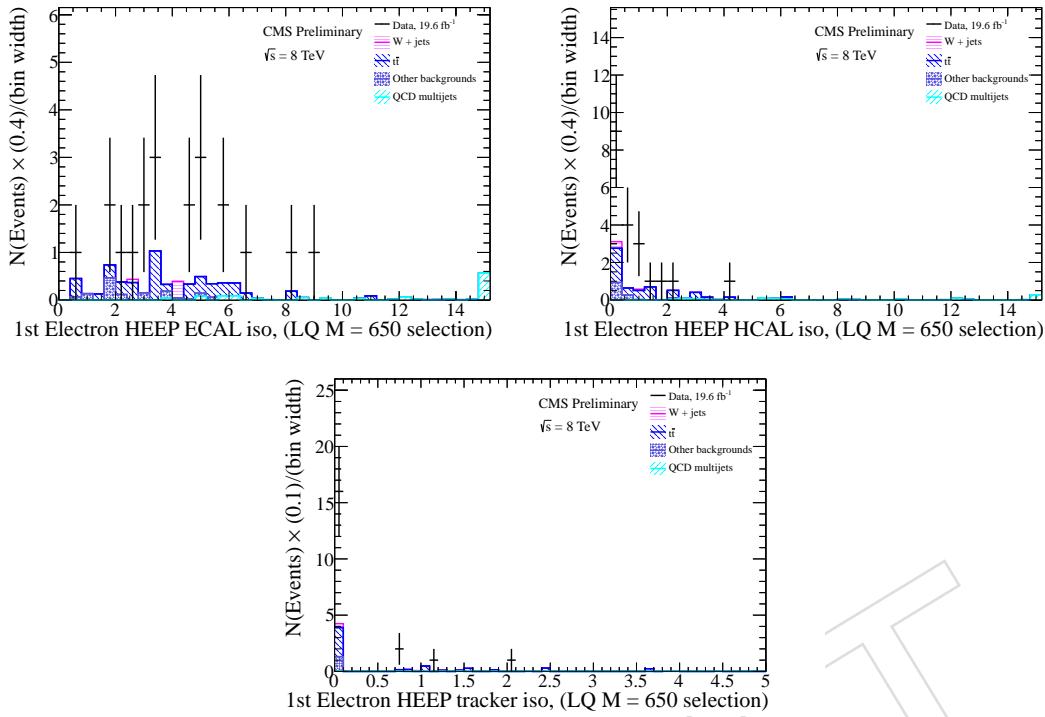


Figure 58: The ECAL isolation (top left), HCAL isolation (top right), and tracker isolation (bottom) for electrons in events passing the  $evjj$  M(LQ) = 650 final selection. Shape discrepancies in these distributions could be a sign of jet contamination in the electron isolation cone.

## 838      **G $evjj$ analysis M(LQ) = 650 final selection invariant mass** 839      **plots**

840      Distributions for the invariant mass of various objects in events passing the M(LQ)  
 841      = 650 final selection.

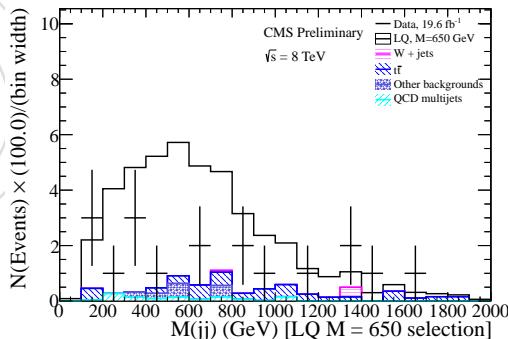


Figure 59: The invariant mass of the two leading jets in events passing the  $evjj$  M(LQ) = 650 final selection.

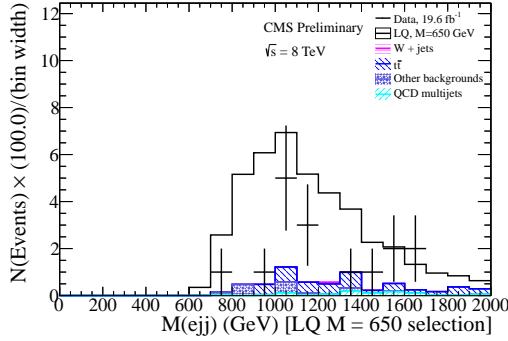


Figure 60: The invariant mass of the two leading jets and the leading electron in events passing the  $evjj$   $M(LQ) = 650$  final selection.

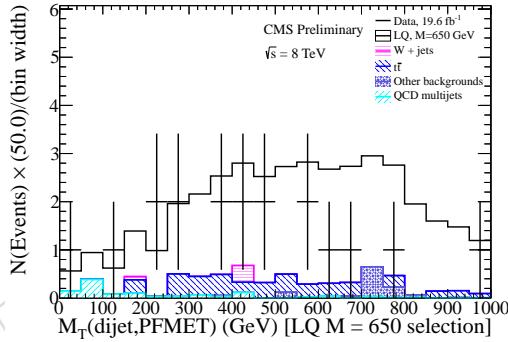


Figure 61: The transverse invariant mass of the two leading jets and the  $E_T^{\text{miss}}$  in events passing the  $evjj$   $M(LQ) = 650$  final selection.

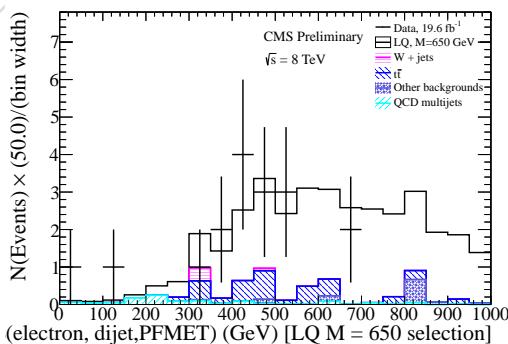


Figure 62: The transverse invariant mass of the electron, two leading jets, and the  $E_T^{\text{miss}}$  in events passing the  $evjj$   $M(LQ) = 650$  final selection.

## H $evjj$ analysis M(LQ) = 650 run period dependence

Distribution showing where in the 2012 run the events contributing to the M(LQ) = 450 and 650 final selections were collected.

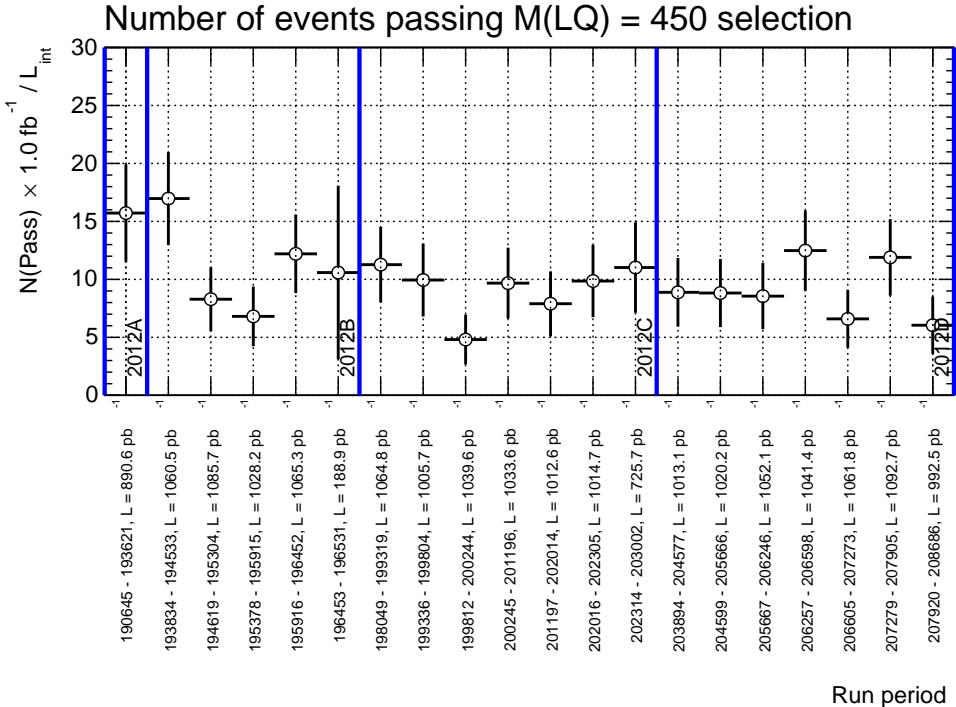


Figure 63: Distribution showing the number of events passing the M(LQ) = 450 final selection of the  $evjj$  analysis during 20 run periods. Each run period corresponds to roughly  $1 \text{ fb}^{-1}$ . The minimum run, maximum run, and integrated luminosity of each run period is listed on the x-axis.

## I Results without systematics

The limits in Figures 65 and the p-values in Figure 66 were calculated using the asymptotic CLs tool from the Higgs group<sup>4</sup>. Unlike the limits in Figure 29, Figure 30, and Figure 28, only statistical uncertainties were included in the limit calculation. Using Figure 65, 95% CL exclusion limits may be placed on first generation scalar leptoquarks with masses less than 1010(850) GeV, assuming  $\beta = 1.0(0.5)$ . This is to be compared with expected 95% CL exclusions of 1030(895) GeV, assuming  $\beta = 1.0(0.5)$ . This deviates very little from the limits calculated with systematic uncertainties in Section 7, which implies that statistical uncertainties dominate.

## J Comparison with LQ2

The second generation leptoquark analysis (EXO-12-042) searches for leptoquarks that decay to leptons (muons or muon neutrinos) and quarks (charm or strange) of the second mass generation. Because the Standard Model backgrounds that contribute significantly to the second generation leptoquark analysis at preselection are slightly different from the Standard Model backgrounds that contribute significantly

<sup>4</sup> Command used for p-values calculation: `combine -M ProfileLikelihood -n <label> -m <lq mass> --signif --pvalue -d <data_card>`

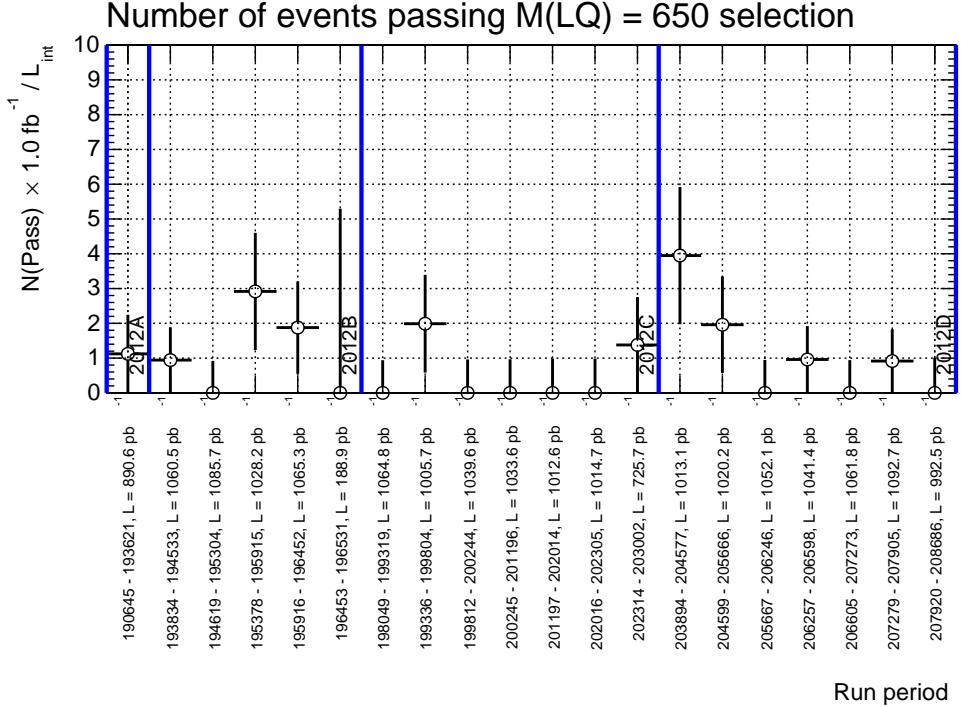


Figure 64: Distribution showing the number of events passing the  $M(LQ) = 650$  final selection of the  $evjj$  analysis during 20 run periods. Each run period corresponds to roughly  $1 \text{ fb}^{-1}$ . The minimum run, maximum run, and integrated luminosity of each run period is listed on the x-axis.

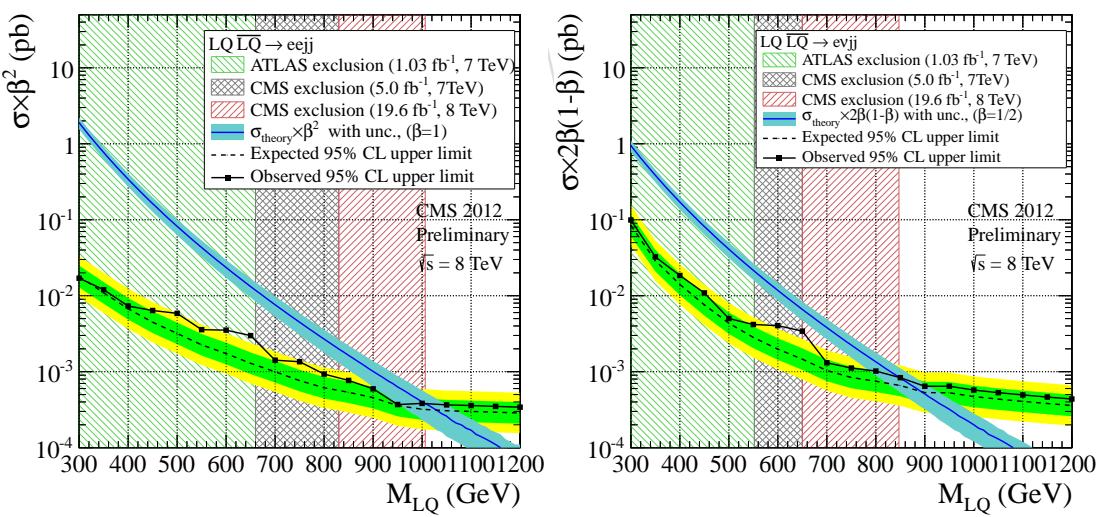


Figure 65: The expected and observed upper limit at 95% C.L. on the LQ pair production cross section times  $\beta^2$  in the top plot ( $2\beta(1 - \beta)$  in the bottom plot) as a function of the LQ mass obtained with the  $eejj$  ( $evjj$ ) analysis. **Only statistical uncertainties are included in the calculation.** The dark blue curve and the light blue band represent, respectively, the theoretical LQ pair production cross section and the uncertainties due to the choice of PDF and renormalization/factorization scales.

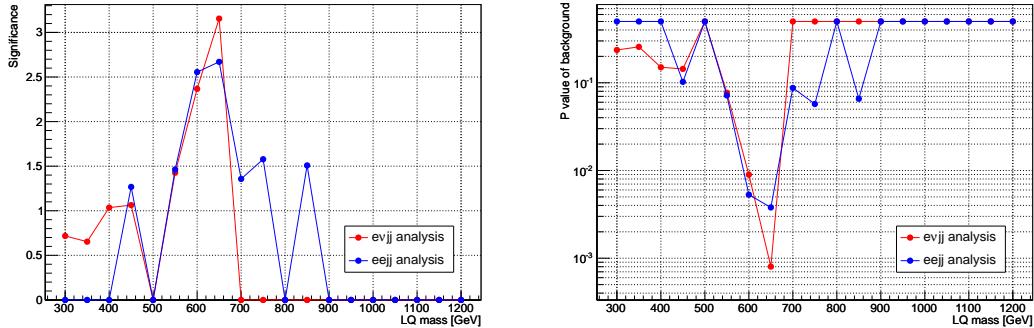


Figure 66: The excess significance (left plot) and p value of the background (right plot) as a function of leptoquark mass.

to this analysis at preselection (namely, the second generation leptoquark analysis does not consider multijet events that include fake leptons), each analysis uses a different final selection.

The second generation leptoquark analysis does not observe a significant excess at  $M(LQ) = 650$  GeV. It is interesting, therefore, to apply the final selection cuts used by the second generation leptoquark analysis to this analysis and to compare the results. This comparison is only possible between the  $eejj$  and the  $\mu\mu jj$  analyses, since the QCD multijet background is negligible for both of these analyses at final selection. The QCD multijet background is not negligible for the  $evjj$  analysis at final selection. The results for both the  $eejj$  analysis and the  $\mu\mu jj$  analysis after applying the final selection cuts for the  $\mu\mu jj$  analysis are shown in Table 32.

Table 32: Comparison between the  $\mu\mu jj$  analysis and the  $eejj$  analysis final selection results, using the cuts in the  $\mu\mu jj$  analysis. Systematic uncertainties are not provided for the  $eejj$  analysis background estimate.

$M_{LQ}$	$eejj$ Total Background	$eejj$ Data	$\mu\mu jj$ Total Background	$\mu\mu jj$ Data
300	$1444.96 \pm 13.65$	1539	$1415 \pm 20 \pm 45$ (syst)	1461
350	$726.71 \pm 9.78$	759	$730 \pm 15 \pm 16$ (syst)	714
400	$399.70 \pm 7.23$	423	$384.8 \pm 10.7 \pm 9.3$ (syst)	394
450	$208.02 \pm 5.18$	235	$205.3 \pm 7.6 \pm 5.5$ (syst)	210
500	$118.74 \pm 4.00$	145	$121.6 \pm 5.7 \pm 4.8$ (syst)	128
550	$71.50 \pm 3.25$	94	$68.1 \pm 4.2 \pm 2.7$ (syst)	75
600	$42.44 \pm 2.40$	67	$44.7 \pm 3.4 \pm 2.0$ (syst)	44
650	$26.99 \pm 1.93$	43	$28 \pm 2.6 \pm 1.3$ (syst)	24
700	$16.42 \pm 1.52$	22	$18.6 \pm 2.2 \pm 1.3$ (syst)	15
750	$10.27 \pm 1.23$	14	$9.32^{+1.29}_{-1.22} \pm 0.87$ (syst)	11
800	$5.08 \pm 0.77$	10	$6.53^{+1.2}_{-1.3} \pm 0.85$ (syst)	9
850	$2.97 \pm 0.54$	4	$3.88^{+1.0}_{-0.92} \pm 0.67$ (syst)	5
900	$1.71 \pm 0.41$	3	$1.47^{+0.81}_{-0.37} \pm 0.43$ (syst)	3
950	$1.04 \pm 0.31$	1	$0.83^{+0.31}_{-0.26} \pm 0.29$ (syst)	1
1000	$0.62 \pm 0.24$	0	$0.383^{+0.894}_{-0.171} \pm 0.031$ (syst)	0
1050	$0.62 \pm 0.24$	0	$0.383^{+0.894}_{-0.171} \pm 0.031$ (syst)	0
1100	$0.62 \pm 0.24$	0	$0.383^{+0.894}_{-0.171} \pm 0.031$ (syst)	0
1150	$0.62 \pm 0.24$	0	$0.383^{+0.894}_{-0.171} \pm 0.031$ (syst)	0
1200	$0.62 \pm 0.24$	0	$0.383^{+0.894}_{-0.171} \pm 0.031$ (syst)	0

Table 32 shows good agreement between the predicted number of  $eejj$  background events, the predicted number of  $\mu\mu jj$  background events, and the number of observed  $\mu\mu jj$  events. Indeed, the number of observed  $eejj$  events is significantly larger than all of these values up to  $M(LQ) = 650$  GeV.

## K Reweighting $E_T^{\text{miss}}$ and $m_{T, e\nu}$ in the $e\nu jj$ analysis

The  $E_T^{\text{miss}}$  and  $m_{T, e\nu}$  distributions at preselection in the  $e\nu jj$  analysis show a discrepancy of roughly 20% between events observed and events predicted. This can be seen in Figure 67.

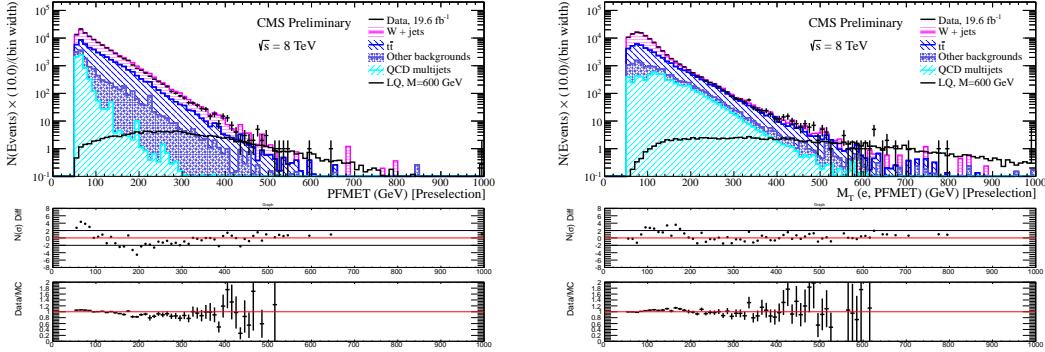


Figure 67: The  $E_T^{\text{miss}}$  (left) and  $m_{T, e\nu}$  (right) distributions for events passing the  $e\nu jj$  preselection.

If one assumes that this discrepancy is caused by a feature in the data that is poorly modeled by the Monte Carlo, it is reasonable to assign a weight to each Monte Carlo event to improve the modeling of these distributions at  $e\nu jj$  preselection.

This appendix describes a procedure by which appropriate weights are calculated and then applied to the Monte Carlo. The procedure is detailed in Section K.1. The results of this reweighting at  $e\nu jj$  preselection are discussed in Section K.2. The results of this reweighting at  $e\nu jj$  final selection are discussed in Section K.3.

### K.1 Procedure

The total weight to be applied to each Monte Carlo event,  $w_{\text{total}}$ , is determined for each event passing the  $e\nu jj$  preselection as a function of that event's values for  $E_T^{\text{miss}}$  and  $m_{T, e\nu}$ .  $w_{\text{total}}$  is evaluated as the product of two separate weights:  $w_1$  (a function of  $E_T^{\text{miss}}$ ) and  $w_2$  (a function of  $m_{T, e\nu}$ ), as described in Equation 13:

$$w_{\text{total}}(E_T, m_{T, e\nu}) = w_1(E_T) \cdot w_2(m_{T, e\nu}) \quad (13)$$

The weights  $w_1$  and  $w_2$  are calculated in series, and  $w_1$  is calculated first. The variable  $\mathcal{R}_1$  is defined using the  $E_T$  histogram shown on the left in Figure 67. For a given bin,  $i$ , in this histogram,  $\mathcal{R}_1$  is defined using Equation 14:

$$\mathcal{R}_1(E_T) = \frac{N_{i,\text{Data}}(E_T) - N_{i,\text{QCD}}(E_T)}{N_{i,W+\text{jets}}(E_T) + N_{i,\text{ff}}(E_T) + N_{i,\text{Other}}(E_T)} \quad (14)$$

where  $N_{i,\text{Data}}(E_T)$  is the number of events observed in data in bin  $i$  of the  $E_T$  histogram,  $N_{i,\text{QCD}}(E_T)$  is the number of QCD events predicted in bin  $i$  of the  $E_T$  histogram,  $N_{i,W+\text{jets}}(E_T)$  is the number of W + jet events predicted in bin  $i$  of the  $E_T$  histogram,  $N_{i,\text{ff}}(E_T)$  is the number of events predicted in bin  $i$  of the  $E_T$  histogram, and  $N_{i,\text{Other}}(E_T)$  is the number of events from other backgrounds predicted in bin  $i$  of the  $E_T$  histogram.

The  $\mathcal{R}_1(E_T)$  distribution is plotted in Figure 68.  $w_1(E_T)$  is defined a first degree polynomial that has been fitted to  $\mathcal{R}_1(E_T)$  and is shown as a violet line in Figure 68. The definition of  $w_1(E_T)$  as a function of  $E_T^{\text{miss}}$  is given in Equation 15:

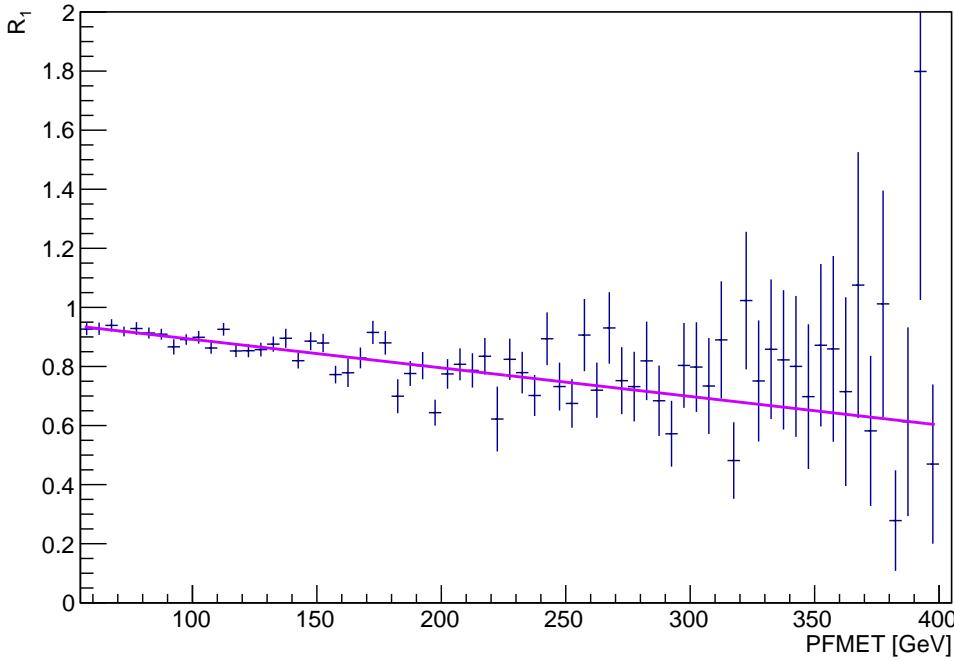


Figure 68:  $\mathcal{R}_1(E_T)$  as defined in Equation 14. The ratio is fit with a first degree polynomial,  $w_1(E_T)$ , which is shown as a violet line.  $w_1(E_T)$  is defined in Equation 15, and its fit parameters are given in Table 33.

Parameter symbol	Parameter title	Mean value	Uncertainty
$a_0$	Linear offset	0.989	$0.0112$
$a_1$	Linear slope	$-9.67 \cdot 10^{-4}$	$8.86 \cdot 10^{-5}$

Table 33: Numeric values and associated uncertainties of the  $w_1(E_T)$  weighting function, defined in Equation 15

$$w_1(E_T) = a_0 + a_1 \cdot E_T \quad (15)$$

893 The numeric values and associated uncertainties of the fit parameters of  $w_1(E_T)$  are  
894 given in Table 33.

Now that  $w_1(E_T)$  has been determined, it is applied to all Monte Carlo events in the analysis. The analysis is repeated with this  $w_1(E_T)$  weight applied, and new scale factors are calculated for W+jets and t̄t Monte Carlo, using the same method described in Section 5.3.2. This yields the following rescaling factors:

$$\begin{aligned} \mathcal{R}_{\bar{t}\bar{t}} &= 1.07 \pm 0.03 \text{ (stat)} \pm 0.01 \text{ (syst)} \\ \mathcal{R}_W &= 0.95 \pm 0.02 \text{ (stat)} \pm 0.01 \text{ (syst)} \end{aligned} \quad (16)$$

895 The new  $E_T$  and  $m_{T, \text{ev}}$  distributions with the  $w_1(E_T)$  weight and the new W+jets and  
896 t̄t scale factors applied to the Monte Carlo is compared with the original distribution  
897 in Figure 69.

898 It is clear from Figure 69 that applying the  $w_1(E_T)$  weight to the Monte Carlo sig-  
899 nificantly improves the agreement between the background estimate and the ob-  
900 served  $E_T$  distribution in data at  $e\nu jj$  preselection. However, the  $m_{T, \text{ev}}$  distribution  
901 at  $e\nu jj$  preselection still exhibits a significant discrepancy after the  $w_1(E_T)$  weight has  
902 been applied.

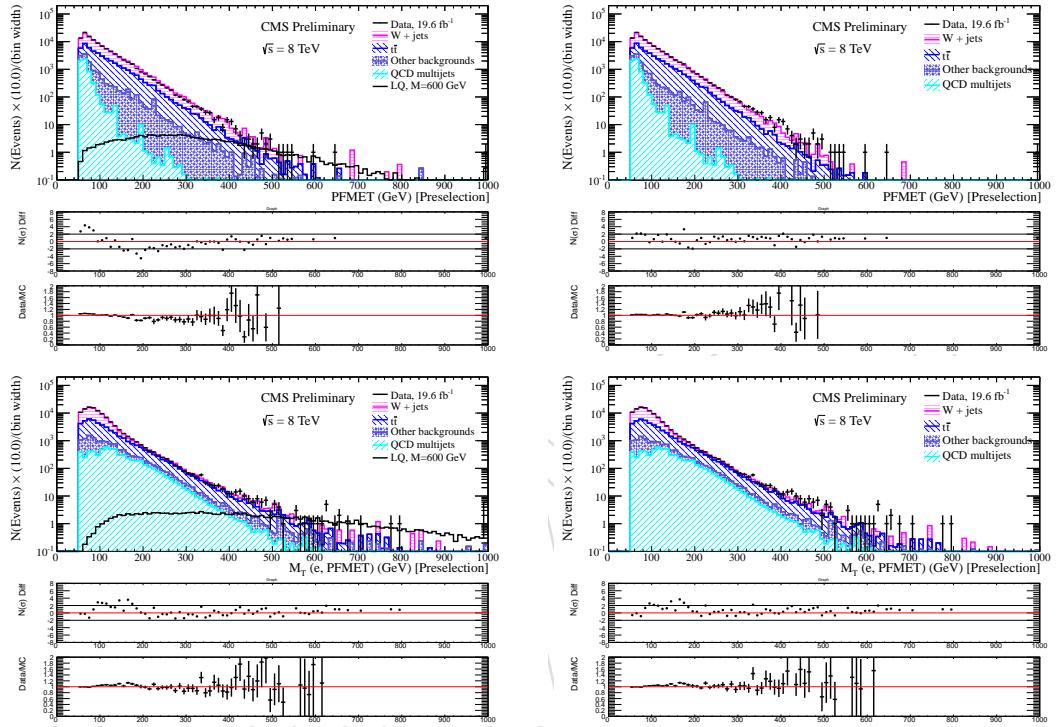


Figure 69: The  $E_T^{\text{miss}}$  distributions at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_1(E_T^{\text{miss}})$  function (top, left) and after being reweighted (top, right). The  $m_{T, e\nu}$  distributions at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_1(E_T^{\text{miss}})$  function (bottom, left) and after being reweighted (bottom, right) are also shown. In the case of the  $E_T^{\text{miss}}$  distribution (top plots), the agreement between the events observed in data and the background estimate is significantly improved by the reweighting. There is little change in the case of the  $m_{T, e\nu}$  distribution.

This discrepancy in the  $m_{T, \text{ev}}$  distribution is dealt with using a similar method to the method for the  $E_T$  distribution described above. A variable  $\mathcal{R}_2$  is defined using the  $m_{T, \text{ev}}$  histogram shown on the bottom-right in Figure 69, before the W+jets and  $t\bar{t}$  rescaling factors have been applied. For a given bin,  $i$ , in this histogram,  $\mathcal{R}_2$  is defined using Equation 17:

$$\mathcal{R}_2(m_{T, \text{ev}}) = \frac{N_{i,\text{Data}}(m_{T, \text{ev}}) - N_{i,\text{QCD}}(m_{T, \text{ev}})}{N_{i,\text{W+jets}}(m_{T, \text{ev}}) + N_{i,\bar{t}\bar{t}}(m_{T, \text{ev}}) + N_{i,\text{Other}}(m_{T, \text{ev}})} \quad (17)$$

where  $N_{i,\text{Data}}(m_{T, \text{ev}})$  is the number of events observed in data in bin  $i$  of the  $m_{T, \text{ev}}$  histogram,  $N_{i,\text{QCD}}(m_{T, \text{ev}})$  is the number of QCD events predicted in bin  $i$  of the  $m_{T, \text{ev}}$  histogram,  $N_{i,\text{W+jets}}(m_{T, \text{ev}})$  is the number of W + jet events predicted in bin  $i$  of the  $m_{T, \text{ev}}$  histogram,  $N_{i,\bar{t}\bar{t}}(m_{T, \text{ev}})$  is the number of events predicted in bin  $i$  of the  $m_{T, \text{ev}}$  histogram, and  $N_{i,\text{Other}}(m_{T, \text{ev}})$  is the number of events from other backgrounds predicted in bin  $i$  of the  $m_{T, \text{ev}}$  histogram.

The  $\mathcal{R}_2(m_{T, \text{ev}})$  distribution is plotted in Figure 70.  $w_2(m_{T, \text{ev}})$  is defined the sum of a first degree polynomial and a gaussian that has been fitted to  $\mathcal{R}_2(m_{T, \text{ev}})$  and is shown as a violet line in Figure 70. The definition of  $w_2(m_{T, \text{ev}})$  as a function of

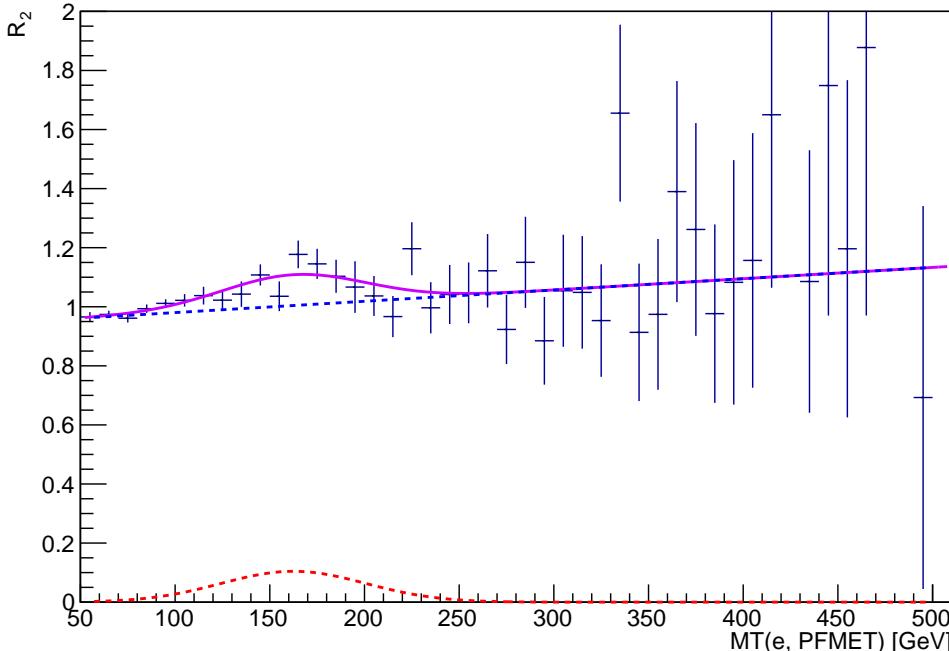


Figure 70:  $\mathcal{R}_2(m_{T, \text{ev}})$  as defined in Equation 17. The ratio is fit with the sum of a first degree polynomial and a gaussian,  $w_2(m_{T, \text{ev}})$ , which is shown as a violet line.  $w_2(m_{T, \text{ev}})$  is defined in Equation 18, and its fit parameters are given in Table 34.

$m_{T, \text{ev}}$  is given in Equation 18.

$$w_2(m_{T, \text{ev}}) = b_0 + b_1 \cdot m_{T, \text{ev}} + B \cdot e^{-\frac{1}{2} \cdot \left(\frac{m_{T, \text{ev}} - \mu}{\sigma}\right)^2} \quad (18)$$

The numeric values and associated uncertainties of the fit parameters of  $w_2(m_{T, \text{ev}})$  are given in Table 34.

Parameter symbol	Parameter title	Mean value	Uncertainty
$b_0$	Linear offset	.942	$0.0181$
$b_1$	Linear slope	$3.82 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$
$B$	Gaussian constant	0.104	0.0279
$\mu$	Gaussian width	38.2	11.6
$\sigma$	Gaussian mean	162	10.1

Table 34: Numeric values and associated uncertainties of the  $w_2(m_{T, ev})$  weighting function, defined in Equation 18

Now that  $w_1(E_T)$  and  $w_2(m_{T, ev})$  weights have been determined, the total weight ( $w_{\text{total}}(E_T, m_{T, ev}) = w_1(E_T) \cdot w_2(m_{T, ev})$ ) is applied to all Monte Carlo events in the analysis. The analysis is repeated with this  $w_{\text{total}}(E_T, m_{T, ev})$  weight applied, and new scale factors are calculated for W+jets and  $t\bar{t}$  Monte Carlo, using the same method described in Section 5.3.2. This yields the following rescaling factors:

$$\begin{aligned} \mathcal{R}_{t\bar{t}} &= 1.08 \pm 0.03 \text{ (stat)} \pm 0.01 \text{ (syst)} \\ \mathcal{R}_W &= 0.97 \pm 0.02 \text{ (stat)} \pm 0.01 \text{ (syst)} \end{aligned} \quad (19)$$

The new  $m_{T, ev}$  distribution with the  $w_{\text{total}}(E_T, m_{T, ev})$  weight and the new W+jets and  $t\bar{t}$  scale factors applied to the Monte Carlo is compared with the original distribution in Figure 71. The improvement in the agreement for both distributions is significant.

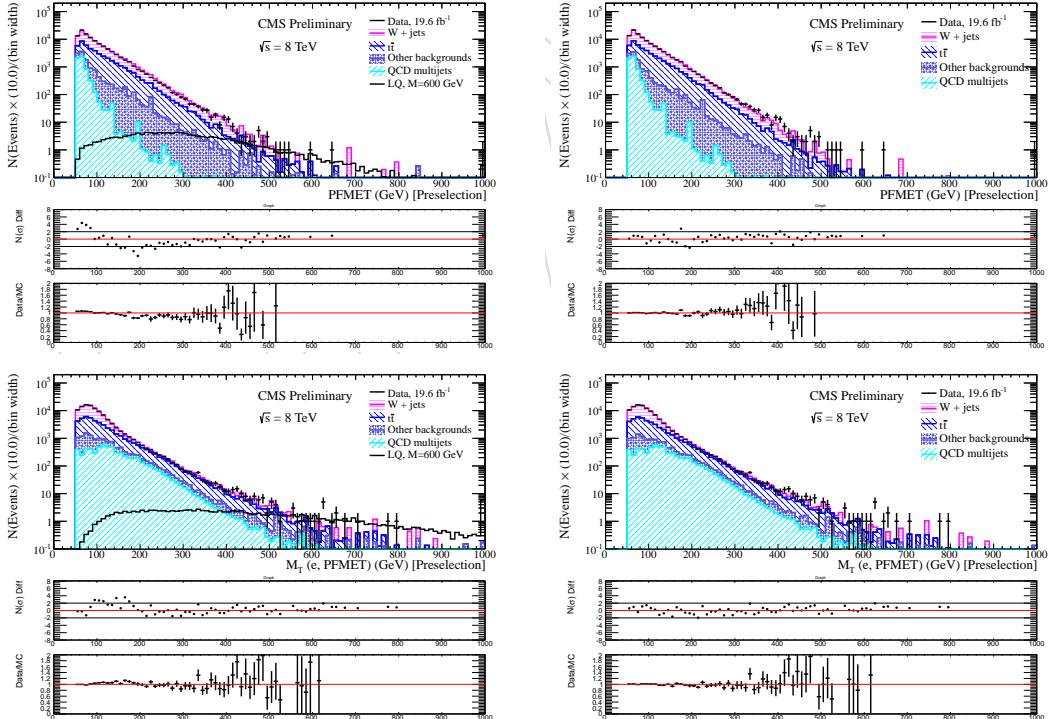


Figure 71: The  $E_T^{\text{miss}}$  distributions at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_{\text{total}}(E_T^{\text{miss}}, m_{T, ev})$  function (top, left) and after being reweighted (top, right). The  $m_{T, ev}$  distributions at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_{\text{total}}(E_T^{\text{miss}}, m_{T, ev})$  function (bottom, left) and after being reweighted (bottom, right) are also shown. The agreement between the events observed in data and the background estimate is significantly improved in both distributions by the reweighting.

## K.2 Results at preselection

Applying the reweighting described in the previous section changes only a few distributions at preselection:

- Reweighting the  $E_T$  and  $m_{T, ev}$  distributions significantly improves the modeling of the  $\Delta\phi(E_T^{\text{miss}}, e)$  distribution at  $evjj$  preselection. This can be seen in Figure 72.
- Reweighting the  $E_T$  distribution moderately improves the low- $S_T$  region of the  $S_T$  distribution, which was affected by poor  $E_T^{\text{miss}}$  modeling. This can be seen in Figure 73.
- Reweighting the  $E_T$  and  $m_{T, ev}$  distributions causes no significant change to the  $m_{ej}$  distribution at  $evjj$  preselection, the only remaining final selection variable in the  $evjj$  analysis.

None of the other preselection distributions change significantly from their original, unweighted descriptions in Section 4.4.1.

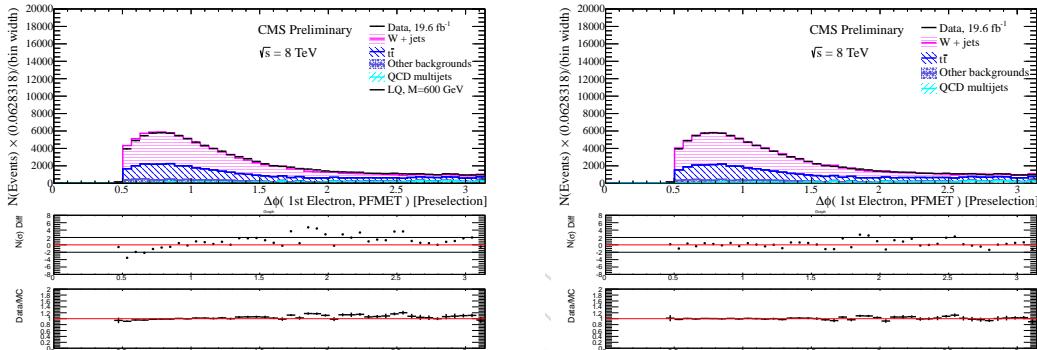


Figure 72: The  $\Delta\phi(E_T^{\text{miss}}, e)$  distribution at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_{\text{total}}(E_T^{\text{miss}}, m_{T, ev})$  function (left) and after being reweighted (right). The agreement between the background estimate and the observed events in data is significantly improved by the reweighting.

## K.3 Results at final selection

The same final selection cuts described in Section 4.4.3 are applied to this reweighted version of the  $evjj$  analysis. The resulting background predictions at each level of final selection are given in Table K.3. Due to the reweighting method described above, the predicted number of events passing the  $M(\text{LQ}) = 650 \text{ GeV}$  selection has decreased from  $7.54 \pm 1.20$  (without reweighting, see Table 13) to  $6.43 \pm 1.00$  (with reweighting, see below). The number of events in observed data is unchanged. The reweighting, therefore, increases the significance of the excess.

## L Purely data-driven background estimate using muons

All of the Standard Model backgrounds in both the  $evjj$  and  $eejj$  analyses may be placed into one of the following two categories:

- Events from Standard Model processes containing isolated high- $p_T$  leptons
- Events from Standard Model processes containing jets which fake isolated high- $p_T$  leptons

The shapes of the distributions of the backgrounds in Category 1 are modeled using Monte Carlo in all cases except for the  $t\bar{t}$  background estimate in the  $eejj$  analysis

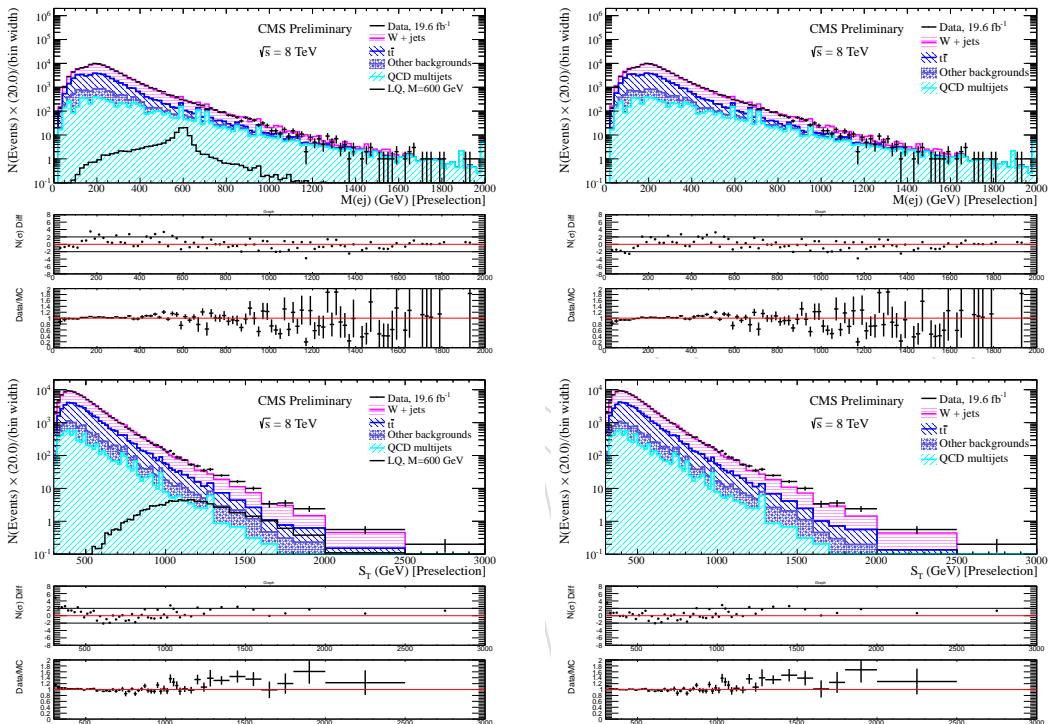


Figure 73: The  $S_T$  and  $m_{e_j}$  distributions at  $evjj$  preselection, before the Monte Carlo has been reweighted with the  $w_{\text{total}}(E_T^{\text{miss}}, m_T, e_V)$  function (left) and after being reweighted (right). There is little change in the  $m_{e_j}$  distribution (top row). The agreement between the background estimate and the observed events in data at the low edge of the  $S_T$  distribution (bottom row) is somewhat improved due to the improvement in the  $E_T^{\text{miss}}$  distribution.

Table 35: Number of events after final  $evjj$  selection, after the  $w_{\text{total}}(E_{\text{T}}^{\text{miss}}, m_{\text{T}, \text{ev}})$  weighting function has been applied to the Monte Carlo. Only statistical errors are reported. Note that the predicted number of events passing the  $M(\text{LQ}) = 650$  GeV selection has decreased from  $7.54 \pm 1.20$  (without reweighting, see Table 13) to  $6.43 \pm 1.00$  (with reweighting, see below). The number of events in observed data is unchanged. The reweighting, therefore, increases the significance of the excess.

$M_{\text{LQ}}$	W+Jets	$t\bar{t}$	QCD	Other	Data	Total BG
Presel	$59725.3 \pm 201.9$	$33176.5 \pm 71.7$	$5950.5 \pm 20.1$	$5943.8 \pm 205.5$	105164	$104796.0 \pm 297.6$
300	$859.6 \pm 23.1$	$1233.0 \pm 12.4$	$117.9 \pm 1.5$	$187.6 \pm 6.907$	2455	$2398.04 \pm 27.16$
350	$280.4 \pm 14.4$	$446.3 \pm 7.3$	$59.11 \pm 0.97$	$88.6 \pm 4.649$	908	$874.28 \pm 16.83$
400	$108.5 \pm 7.7$	$180.6 \pm 4.6$	$32.88 \pm 0.69$	$43.7 \pm 3.229$	413	$365.71 \pm 9.55$
450	$50.5 \pm 5.5$	$70.8 \pm 2.8$	$14.13 \pm 0.42$	$21.3 \pm 2.257$	192	$156.75 \pm 6.62$
500	$19.0 \pm 3.0$	$31.6 \pm 1.9$	$7.76 \pm 0.30$	$12.4 \pm 1.734$	83	$70.81 \pm 3.93$
550	$7.9 \pm 1.7$	$13.3 \pm 1.2$	$3.89 \pm 0.21$	$6.3 \pm 1.288$	44	$31.36 \pm 2.43$
600	$2.2 \pm 0.9$	$6.13 \pm 0.80$	$2.29 \pm 0.17$	$3.5 \pm 0.959$	28	$14.08 \pm 1.57$
650	$0.43 \pm 0.30$	$3.22 \pm 0.59$	$1.18 \pm 0.12$	$1.59 \pm 0.736$	18	$6.43 \pm 1.00$
700	$0.43 \pm 0.30$	$2.17 \pm 0.50$	$0.85 \pm 0.10$	$0.35 \pm 0.150$	6	$3.80 \pm 0.61$
750	$0.00^{+0.94}_{-0.00}$	$1.49 \pm 0.41$	$0.514 \pm 0.091$	$0.117 \pm 0.061$	4	$2.116^{+1.025}_{-0.420}$
800	$0.00^{+0.94}_{-0.00}$	$0.87 \pm 0.30$	$0.317 \pm 0.067$	$0.116 \pm 0.061$	3	$1.308^{+0.986}_{-0.313}$
850	$0.00^{+0.94}_{-0.00}$	$0.70 \pm 0.27$	$0.117 \pm 0.029$	$0.054 \pm 0.032$	2	$0.874^{+0.975}_{-0.278}$
900	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.076 \pm 0.024$	$0.019 \pm 0.012$	1	$0.366^{+0.948}_{-0.159}$
950	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$
1000	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$
1050	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$
1100	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$
1150	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$
1200	$0.00^{+0.94}_{-0.00}$	$0.27 \pm 0.16$	$0.069 \pm 0.023$	$0.019 \pm 0.012$	1	$0.359^{+0.948}_{-0.159}$

(see Section 5.3.1). The shapes of the distributions of the backgrounds in Category 2 are modeled using data (see Section 5.2). In order to determine whether poor Monte Carlo modeling of events in Category 1 is behind the excess observed at final selection in both analyses, this section describes a crosscheck whereby each analysis is repeated after modeling Category 1 backgrounds using muon events from data.

In this method,  $\mu\mu jj$  events are used to model *eejj* events (in the *eejj* analysis), and  $\mu\nu jj$  events are used to model *evjj* events (in the *evjj* analysis). In both analyses, isolated muons are treated as electrons, and each muon event is weighted to account for differences in reconstruction and trigger efficiencies for electrons and muons. The fake rate method described in Section 5.2 remains unchanged from the main analysis. This method of estimating the contribution of Category 1 backgrounds excludes any contamination from first generation leptoquark signal, because first generation leptoquarks are forbidden to decay to second generation leptons (muons).

In the case of the *eejj* analysis, Section L.1 of this appendix describes the method for rescaling  $\mu\mu jj$  events to model *eejj* events. Section L.2 describes the results of this method at *eejj* preselection. Section L.3 describes the results of this method at *eejj* final selection. Similarly, in the case of the *evjj* analysis, Section L.4 of this appendix describes the method for rescaling  $\mu\mu jj$  events to model *evjj* events. Section L.5 describes the results of this method at *evjj* preselection. Section L.6 describes the results of this method at *evjj* final selection. Section L.7 draws conclusions from this cross-check.

## L.1 Procedure for the *eejj* analysis

As discussed above, the shapes of the predicted background distributions in the *eejj* analysis are taken from a control sample containing  $\mu\mu jj$  events. The *eejj* preselection and final selection requirements described in Section 4.3 are applied to the  $\mu\mu jj$  sample as well, where the muon is treated as an electron. These muon events are selected online using the same unprescaled single muon triggers described in Table 17 of Section 5.3.1.

The normalization of the the predicted background distributions in the *eejj* analysis are taken by weighting each muon event according to Equation 20:

$$N_{eejj}^{\text{data}} = C_{\mu\mu jj} \times N_{\mu\mu jj}^{\text{data}} = \left( \frac{\epsilon_{eejj}^{\text{trigger}}}{\epsilon_{\mu}^{\text{trigger}}} \times \frac{\epsilon_{eejj}^{\text{reco}/\text{ID}/\text{Iso}}}{\epsilon_{\mu\mu jj}^{\text{reco}/\text{ID}/\text{Iso}}} \right) \times N_{\mu\mu jj}^{\text{data}} \quad (20)$$

where:

- $\epsilon_{eejj}^{\text{trigger}}$  ( $\epsilon_{\mu}^{\text{trigger}}$ ) is the trigger efficiency to select single electron-dijet (single muon) events with the HLT paths listed in Tables 8 and 9 (Table 17). As discussed in Section 4.2, the efficiency ( $\epsilon_{eejj}^{\text{trigger}}$ ) of the single electron-dijet trigger on events with at least one HEEP electron and two high- $p_T$  jets is  $97.4 \pm 0.56\%$  for electrons in the barrel and  $95.8 \pm 1.35\%$  for electrons in the endcap [30]. The efficiency ( $\epsilon_{\mu}^{\text{trigger}}$ ) of the single muon trigger on muons passing the muon ID described in Section 3.2 is binned in muon  $|\eta|$  and is taken to be 0.94 (for  $|\eta| \leq 0.9$ ), 0.84 (for  $0.9 < |\eta| \leq 1.2$ ), 0.82 (for  $1.2 < |\eta| \leq 2.1$ ). To account for statistical uncertainties and variation over different data taking periods, a conservative uncertainty of 1% is placed on these efficiencies.

- $\epsilon_{eejj}^{\text{reco}/\text{ID}/\text{Iso}}$  ( $\epsilon_{\mu\mu jj}^{\text{reco}/\text{ID}/\text{Iso}}$ ) is the  $eejj$  ( $\mu\mu jj$ ) reconstruction, ID, and isolation efficiency events passing the above triggers. The efficiency ratio is taken from Monte Carlo  $eejj$  preselection level:

$$\epsilon_{eejj}^{\text{reco}/\text{ID}/\text{Iso}} / \epsilon_{\mu\mu jj}^{\text{reco}/\text{ID}/\text{Iso}} = N_{eejj}^{\text{MC}} / N_{\mu\mu jj}^{\text{MC}} = 97.5 \pm 0.4 \text{ (stat)} \quad (21)$$

984 In the case of the  $eejj$  analysis, the difference between the muon and electron energy  
 985 resolutions leads to a discrepancy between the shape of the  $Z \rightarrow \ell\ell$  mass peak pre-  
 986 dicted by the muon events and the shape observed in electron events. This can be  
 987 be seen in Figure 74. A smaller but similar discrepancy exists in the shape of the  $Z \rightarrow \ell\ell$   
 988 mass peak as predicted by Monte Carlo and the shape observed in electron events in  
 989 data. Both of these discrepancies may be seen in Figure 74. As in the main analysis,  
 990 no effort is made to correct this discrepancy, since the agreement in the tail of the  
 991 dilepton mass distribution is relatively good. This is shown in Figure 75.

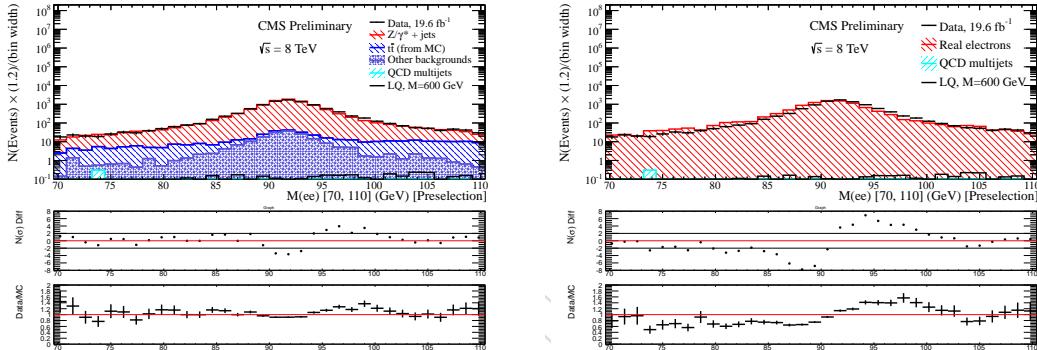


Figure 74: The shape of the  $Z \rightarrow \ell\ell$  mass peak observed in electron events in data as compared to a Monte Carlo background estimate (left) and as compared to a data-driven background estimate using muons (right). A discrepancy exists for both background estimates, due to difference between the predicted and observed lepton energy resolutions.

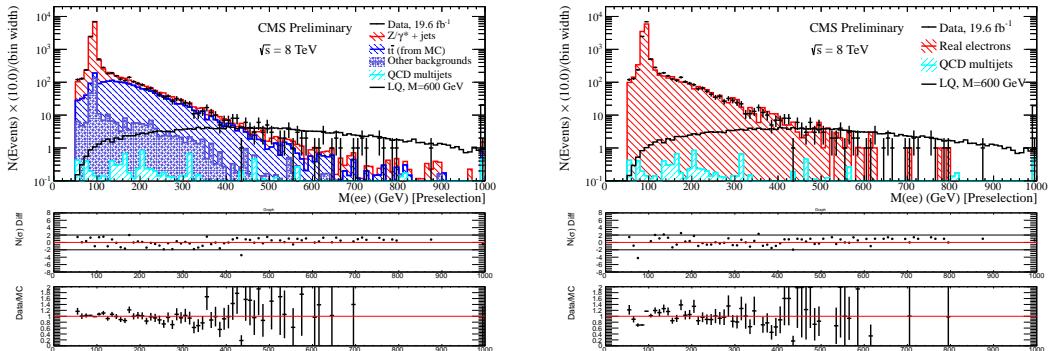


Figure 75: The shape of the  $Z \rightarrow \ell\ell$  mass tails observed in electron events in data as compared to a Monte Carlo background estimate (left) and as compared to a data-driven background estimate using muons (right). Outside of the mass peak region, the agreement between both background estimates and the observed distribution in electron events in data is relatively good.

## L.2 Results at $eejj$ preselection

After the procedure described in Section L.1 has been completed, the preselection of the  $eejj$  analysis as described in Section 4.3.1 is applied. This section shows the

995 agreement between the electron events observed in data and the background esti-  
 996 mate.

997 The distribution of the number of reconstructed primary vertices is shown in Figure  
 998 76. The  $p_T$  and  $\eta$  distributions of the two electrons and the two leading jets are  
 999 shown in Figures 77, 78, 79, and 80. The  $m_{ee}$  distribution is shown in Figure 81. The  
 1000  $S_T$  and  $m_{ej}$  distributions are shown in Figure 82.

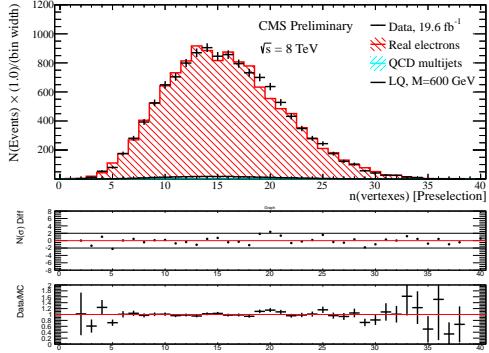


Figure 76: The distribution of the number of primary vertices for events passing the  $eejj$  pre-selection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

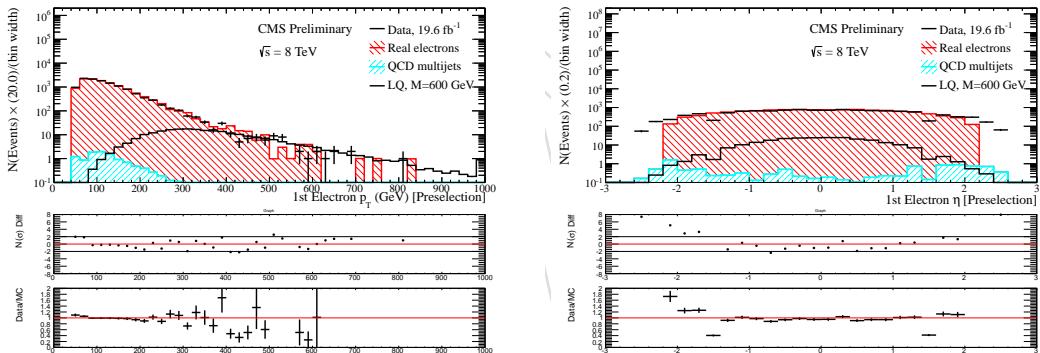


Figure 77: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue). The features in the  $\eta$  (right) distribution stem from the differences in acceptance between the ECAL and the tracker+muon system.

### L.3 Results at $eejj$ final selection

1002 The same final selection cuts described in Section 4.3.3 are applied to this version of  
 1003 the  $eejj$  analysis. The resulting background predictions at each level of final selection  
 1004 are given in Table 36.

1005 It is notable that this background estimate predicts  $26.67 \pm 5.12$  (stat) events at the fi-  
 1006 nal selection optimized for leptoquarks with a mass of 650 GeV. This is in agreement  
 1007 with 20.492.14 (stat) background events predicted by the main analysis. However,  
 1008 because the central value of the data-driven background prediction is higher, and be-  
 1009 cause the statistical uncertainty of the data-driven background prediction is larger,  
 1010 the significance of the excess at final selection for leptoquarks with a mass of 650  
 1011 GeV is less significant under the data-driven background prediction. This can be  
 1012 observed in the limit on leptoquark masses, which is shown in Figure 94.

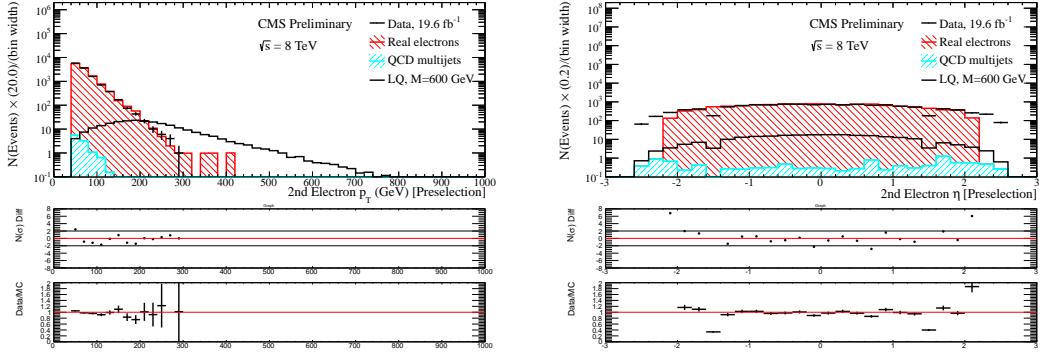


Figure 78: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) electron for events passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue). The features in the  $\eta$  (right) distribution stem from the differences in acceptance between the ECAL and the tracker+muon system.

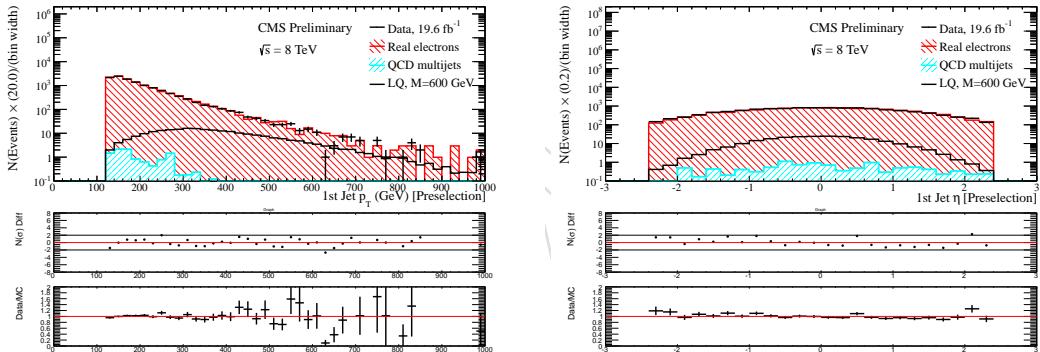


Figure 79: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

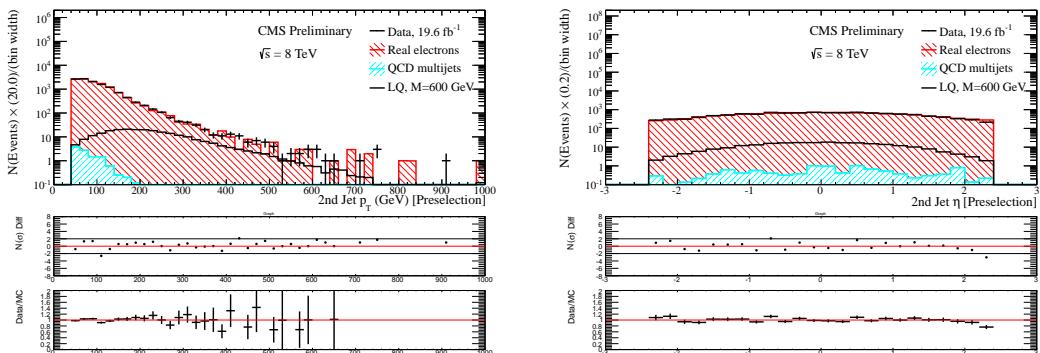


Figure 80: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

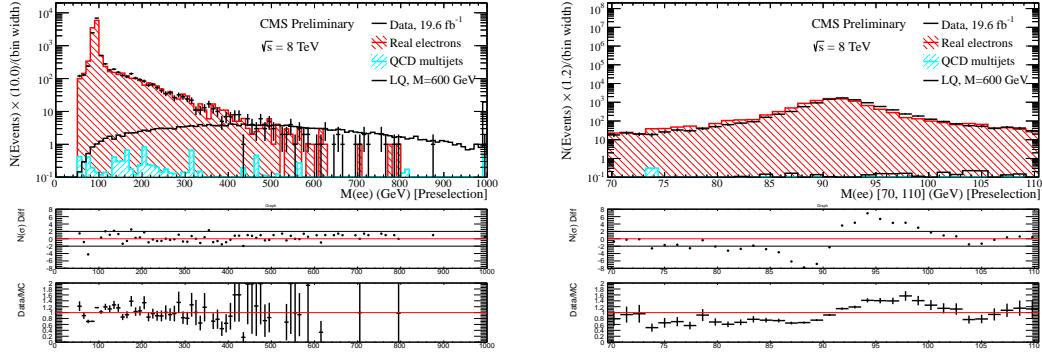


Figure 81: The  $m_{ee}$  distribution in the full range (left) and in a zoomed region around the  $Z^0$  mass peak (right) passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

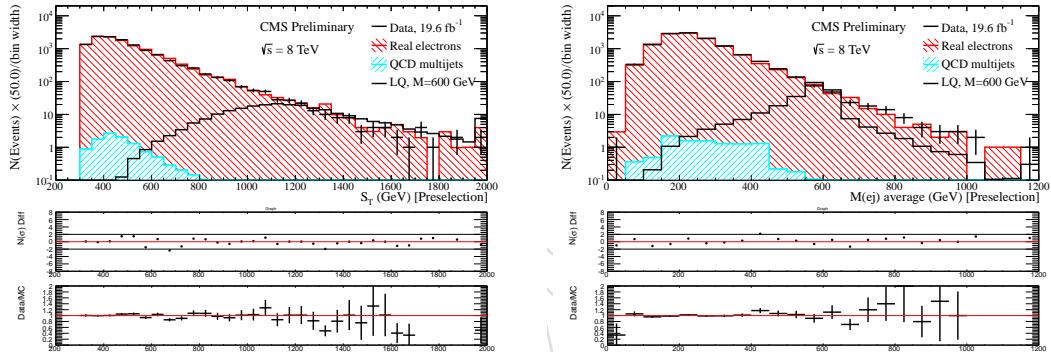


Figure 82: The  $S_T$ (left) and  $m_{ej}$ (right) distributions for events passing the  $eejj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

Table 36: Number of events passing the final  $eejj$  selection, using the purely data-driven muon background estimate. Only statistical errors are reported.

$M_{LQ}$	LQ Signal	Category 1 bkgd.	Category 2 bkgd.	Data	Total BG
Presel	-	$12399.1 \pm 110.7$	$10.87 \pm 0.10$	12442	$12410.0 \pm 110.7$
300	$12855.1 \pm 75.9$	$1146.7 \pm 33.6$	$5.282 \pm 0.052$	1244	$1152.02 \pm 33.63$
350	$6137.3 \pm 31.6$	$677.3 \pm 25.8$	$3.215 \pm 0.036$	736	$680.54 \pm 25.84$
400	$2928.6 \pm 14.2$	$353.0 \pm 18.7$	$1.696 \pm 0.023$	389	$354.66 \pm 18.65$
450	$1429.7 \pm 6.8$	$201.4 \pm 14.1$	$0.890 \pm 0.016$	233	$202.24 \pm 14.10$
500	$727.5 \pm 3.4$	$126.3 \pm 11.2$	$0.485 \pm 0.011$	148	$126.78 \pm 11.16$
550	$389.2 \pm 1.8$	$70.0 \pm 8.3$	$0.2758 \pm 0.0084$	81	$70.25 \pm 8.30$
600	$213.96 \pm 0.98$	$43.4 \pm 6.5$	$0.1527 \pm 0.0065$	57	$43.56 \pm 6.54$
650	$119.31 \pm 0.55$	$26.6 \pm 5.1$	$0.0760 \pm 0.0040$	36	$26.67 \pm 5.12$
700	$69.09 \pm 0.32$	$16.7 \pm 4.1$	$0.0448 \pm 0.0029$	17	$16.77 \pm 4.06$
750	$40.86 \pm 0.19$	$10.8 \pm 3.3$	$0.0258 \pm 0.0023$	12	$10.85 \pm 3.26$
800	$24.81 \pm 0.11$	$8.8 \pm 2.9$	$0.0193 \pm 0.0022$	7	$8.85 \pm 2.94$
850	$15.147 \pm 0.068$	$5.9 \pm 2.4$	$0.0111 \pm 0.0015$	5	$5.89 \pm 2.40$
900	$9.303 \pm 0.042$	$4.9 \pm 2.2$	$0.0069 \pm 0.0012$	3	$4.91 \pm 2.19$
950	$5.770 \pm 0.026$	$4.9 \pm 2.2$	$0.00451 \pm 0.00085$	1	$4.90 \pm 2.19$
1000	$3.659 \pm 0.017$	$2.0 \pm 1.4$	$0.00374 \pm 0.00082$	1	$1.97 \pm 1.39$
1050	$2.442 \pm 0.011$	$2.0 \pm 1.4$	$0.00374 \pm 0.00082$	1	$1.97 \pm 1.39$
1100	$1.6055 \pm 0.0068$	$2.0 \pm 1.4$	$0.00374 \pm 0.00082$	1	$1.97 \pm 1.39$
1150	$1.0686 \pm 0.0044$	$2.0 \pm 1.4$	$0.00374 \pm 0.00082$	1	$1.97 \pm 1.39$
1200	$0.7108 \pm 0.0029$	$2.0 \pm 1.4$	$0.00374 \pm 0.00082$	1	$1.97 \pm 1.39$

#### L.4 Procedure for the $evjj$ analysis

As discussed above, the shapes of the predicted background distributions in the  $evjj$  analysis are taken from a control sample containing  $\mu jj$  events. The  $evjj$  preselection and final selection requirements described in Section 4.4 are applied to the  $\mu jj$  sample as well, where the muon is treated as an electron. As in the  $eejj$  analysis, these muon events are selected online using the same unprescaled single muon triggers described in Table 17 of Section 5.3.1.

The normalization of the the predicted background distributions in the  $eejj$  analysis are taken by weighting each muon event according to Equation 22:

$$N_{evjj}^{\text{data}} = \mathcal{C}_{\mu jj} \times N_{\mu jj}^{\text{data}} = \left( \frac{\epsilon_{e jj}^{\text{trigger}}}{\epsilon_{\mu}^{\text{trigger}}} \times \frac{\epsilon_{evjj}^{\text{reco/ID/Iso}}}{\epsilon_{\mu jj}^{\text{reco/ID/Iso}}} \right) \times N_{\mu jj}^{\text{data}} \quad (22)$$

where:

- $\epsilon_{e jj}^{\text{trigger}}$  ( $\epsilon_{\mu}^{\text{trigger}}$ ) is the trigger efficiency to select single electron-dijet (single muon) events with the HLT paths listed in Tables 8 and 9 (Table 17). These efficiencies are the same as those discussed in the previous section on the procedure for the muon background estimate in the  $eejj$  analysis (Section L.1).
- $\epsilon_{evjj}^{\text{reco/ID/Iso}}$  ( $\epsilon_{\mu jj}^{\text{reco/ID/Iso}}$ ) is the  $evjj$  ( $\mu jj$ ) reconstruction, ID, and isolation efficiency events passing the above triggers. The efficiency ratio is taken from Monte Carlo  $evjj$  preselection level:

$$\epsilon_{evjj}^{\text{reco/ID/Iso}} / \epsilon_{\mu jj}^{\text{reco/ID/Iso}} = N_{evjj}^{\text{MC}} / N_{\mu jj}^{\text{MC}} = 97.2 \pm 0.5 \text{ (stat)} \quad (23)$$

In the case of the  $evjj$  analysis, the difference between the muon and electron energy resolutions leads to a discrepancy between the shape of the  $m_{T, ev}$  peak predicted by the muon events and the shape observed in electron events. This can be seen in Figure 83.

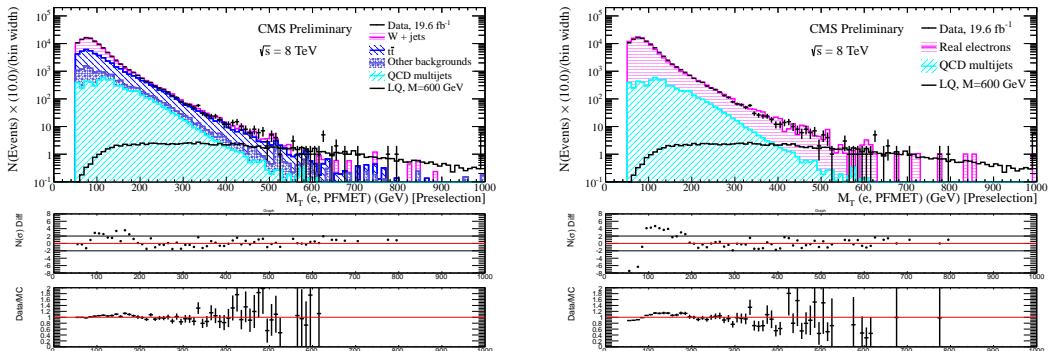


Figure 83: The shape of the  $m_{T, ev}$  peak observed in electron events in data as compared to a Monte Carlo background estimate (left) and as compared to a data-driven background estimate using muons (right). A discrepancy exists for both background estimates, due to difference between the predicted and observed lepton energy resolutions.

Unlike the discrepancy between the predicted and observed shape of the  $Z \rightarrow \ell\ell$  in the  $eejj$  analysis, this discrepancy is significant in the tail of the  $m_{T, ev}$  distribution. As a result, the  $m_{T, ev}$  distribution is reweighted using a method similar to the one described in Appendix K.

As in Appendix K, a variable  $\mathcal{R}_\mu$  is defined using the  $m_{T, ev}$  histogram shown on the right in Figure 83. For a given bin,  $i$ , in this histogram,  $\mathcal{R}_\mu$  is defined using Equation 24:

$$\mathcal{R}_\mu(m_{T, ev}) = \frac{N_{i, \text{Data}}(m_{T, ev}) - N_{i, \text{QCD}}(m_{T, ev})}{N_{i, \mu \text{ background estimate}}(m_{T, ev})} \quad (24)$$

where  $N_{i, \text{Data}}(m_{T, ev})$  is the number of events observed in data in bin  $i$  of the  $m_{T, ev}$  histogram,  $N_{i, \text{QCD}}(m_{T, ev})$  is the number of QCD events predicted in bin  $i$  of the  $m_{T, ev}$  histogram, and  $N_{i, \mu \text{ background estimate}}(m_{T, ev})$  is the number of Category 1 background events predicted in bin  $i$  of the  $m_{T, ev}$  histogram.

The  $\mathcal{R}_\mu(m_{T, ev})$  distribution is plotted in Figure 84. A weight,  $w_\mu(m_{T, ev})$ , is defined as the sum of a first degree polynomial and a gaussian that has been fitted to  $\mathcal{R}_\mu(m_{T, ev})$  and is shown as a violet line in Figure 84. The definition of  $w_\mu(m_{T, ev})$  as a function

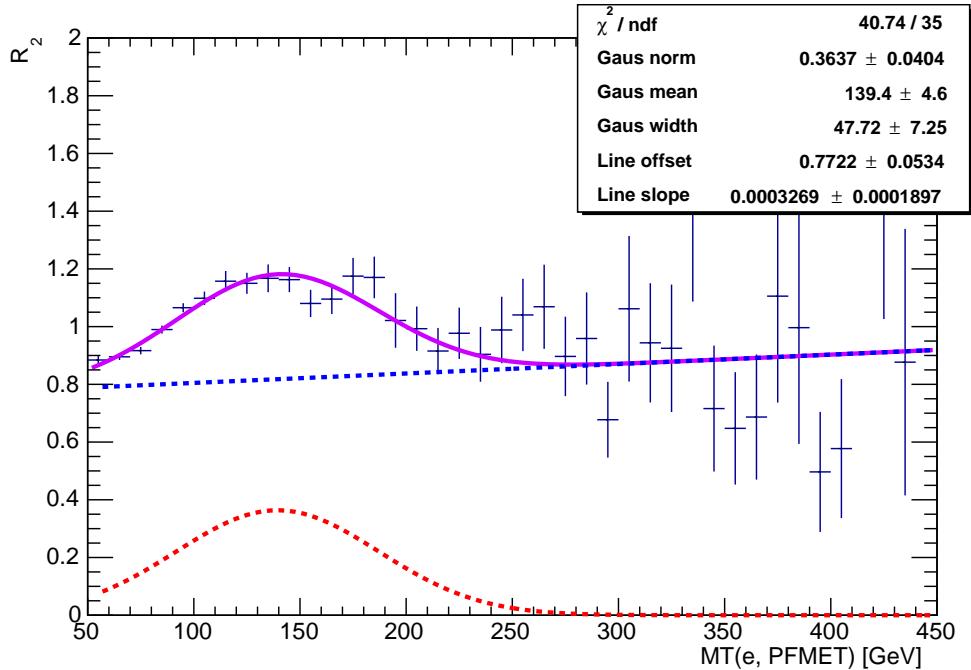


Figure 84:  $\mathcal{R}_\mu(m_{T, ev})$  as defined in Equation 24. The ratio is fit with the sum of a first degree polynomial and a gaussian,  $w_\mu(m_{T, ev})$ , which is shown as a violet line.  $w_\mu(m_{T, ev})$  is defined in Equation 25, and its fit parameters are given in Table 37.

of  $m_{T, ev}$  is given in Equation 25:

$$w_\mu(m_{T, ev}) = b_0 + b_1 \cdot m_{T, ev} + B \cdot e^{-\frac{1}{2} \cdot \left(\frac{m_{T, ev} - \mu}{\sigma}\right)^2} \quad (25)$$

The numeric values and associated uncertainties of the fit parameters of  $w_\mu(m_{T, ev})$  are given in Table 37.

Now that  $w_\mu(m_{T, ev})$  weight has been determined, the analysis is repeated, and  $w_\mu(m_{T, ev})$  is applied to all of the  $\mu\nu jj$  events in the background prediction. The result of this reweighting is shown in Figure 85.

## L.5 Results at $evjj$ preselection

After the procedure described in Section L.4 has been completed, the preselection of the  $evjj$  analysis as described in Section 4.4.1 is applied. This section shows the

Parameter symbol	Parameter title	Mean value	Uncertainty
$b_0$	Linear offset	.772	0.0534
$b_1$	Linear slope	$3.27 \cdot 10^{-4}$	$1.90 \cdot 10^{-4}$
$B$	Gaussian constant	0.363	0.0404
$\mu$	Gaussian width	47.7	7.25
$\sigma$	Gaussian mean	139	4.6

Table 37: Numeric values and associated uncertainties of the  $w_\mu(m_{T, ev})$  weighting function, defined in Equation 25

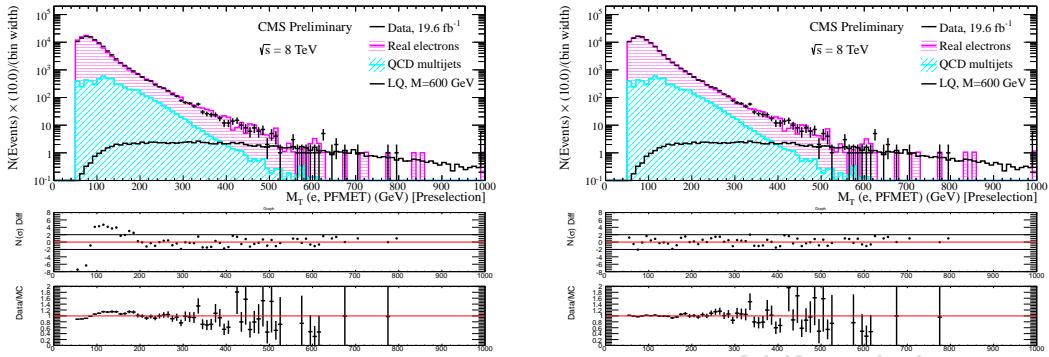


Figure 85: The  $m_{T, ev}$  distributions at  $evjj$  preselection, before the  $\mu vjj$  background estimate events have been reweighted with the  $w_\mu(m_{T, ev})$  function (left) and after being reweighted (right) are shown. The agreement between the events observed in data and the background estimate is significantly improved by the reweighting.

agreement between the electron events observed in data and the background estimate.

The distribution of the number of reconstructed primary vertices is shown in Figure 86. The  $p_T$  and  $\eta$  distributions of the electron and the two leading jets are shown in Figures 87, 88, and 89. The  $E_T^{\text{miss}}$  distribution is shown in Figure 90. The  $\Delta\phi(E_T^{\text{miss}}, e)$ ,  $\Delta\phi(E_T^{\text{miss}}, j1)$ ,  $\Delta\phi(E_T^{\text{miss}}, j2)$ , and  $\text{min}\Delta R(e, \text{jets})$  distributions are shown in Figure 91. The  $m_{T, ev}$  and  $S_T$  distributions are shown in Figure 92. The  $m_{ej}$  and  $m_{Tvj}$  distributions are shown in Figure 93.

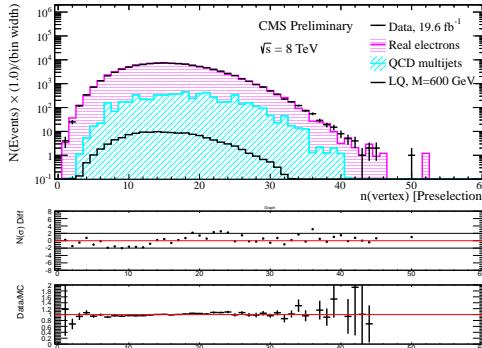


Figure 86: The distribution of the number of primary vertices for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

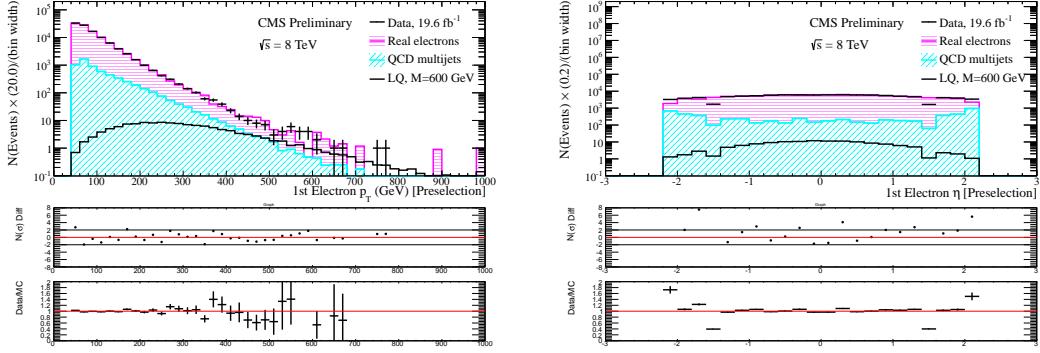


Figure 87: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) electron for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue). The features in the  $\eta$  (right) distribution stem from the differences in acceptance between the ECAL and the tracker+muon system.

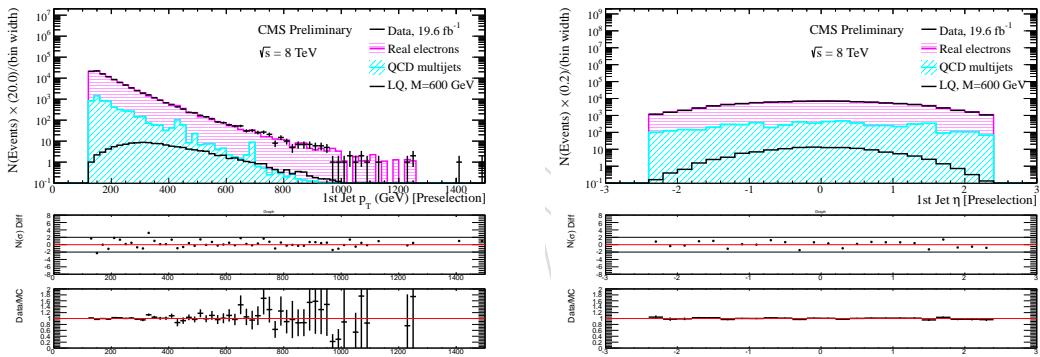


Figure 88: The  $p_T$  (left) and  $\eta$  (right) distributions of the leading (in  $p_T$ ) jet for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

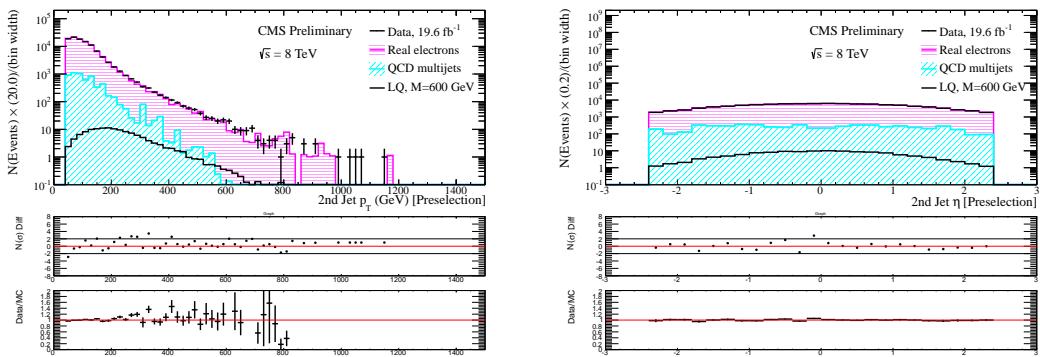


Figure 89: The  $p_T$  (left) and  $\eta$  (right) distributions of the second leading (in  $p_T$ ) jet for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

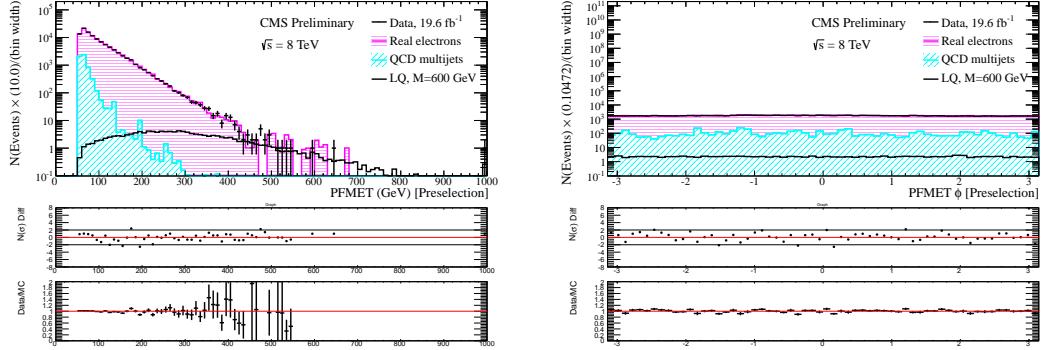


Figure 90: The  $E_{\text{T}}^{\text{miss}}$  (left) and  $\phi(E_{\text{T}}^{\text{miss}})$  (right) distributions for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

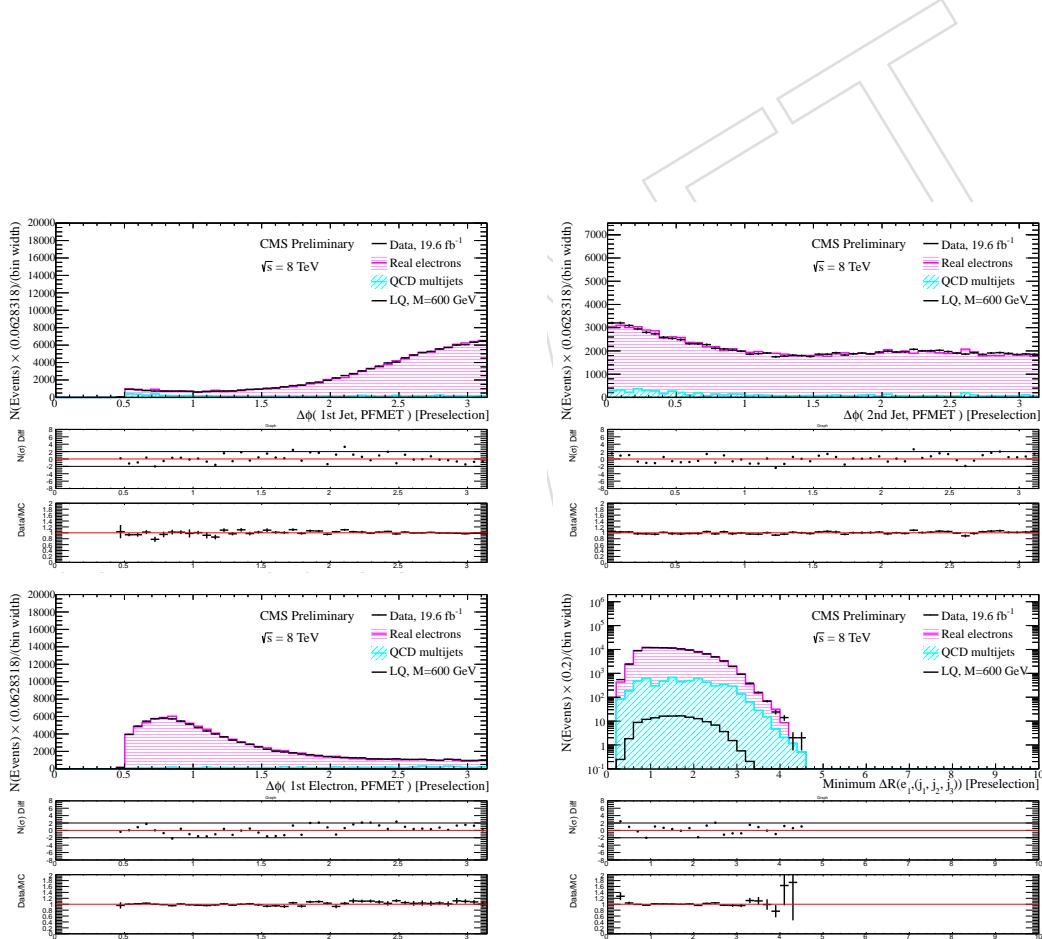


Figure 91: The distribution of  $\Delta|\phi(E_{\text{T}}^{\text{miss}}, \text{j1})$  (top left),  $\Delta|\phi(E_{\text{T}}^{\text{miss}}, \text{j2})$  (top right),  $\Delta|\phi(E_{\text{T}}^{\text{miss}}, e)$  (bottom left), and  $\text{min}\Delta R(e, \text{jets})$  (bottom right) for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

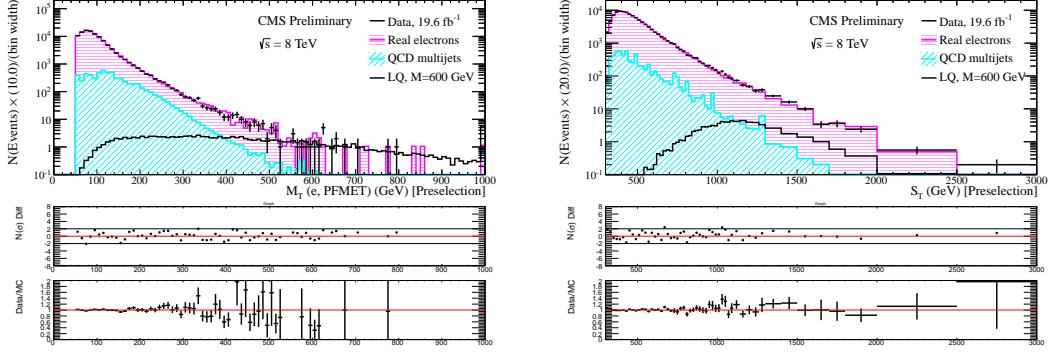


Figure 92: The  $m_{T,\text{e}\nu}$ (left) and  $S_T$ (right) distributions for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

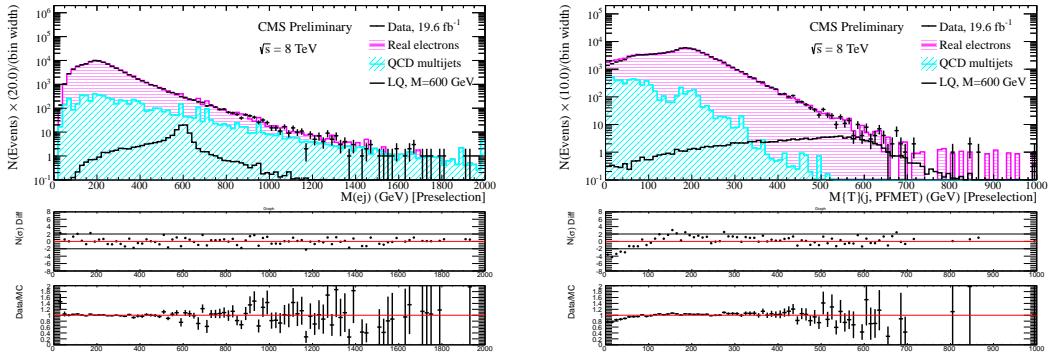


Figure 93: The  $m_{\text{ej}}$ (left) and  $m_{\text{T}jj}$  (right) distributions for events passing the  $evjj$  preselection. The background estimate is derived entirely from data using muon events (red) and the QCD fake rate method (blue).

Table 38: Number of events passing the final  $evjj$  selection, using the purely data-driven muon background estimate. Only statistical errors are reported.

$M_{LQ}$	LQ Signal	Category 1 bkgd.	Category 2 bkgd.	Data	Total BG
Presel	-	$99103.1 \pm 323.9$	$5950.5 \pm 20.1$	105164	$105053.6 \pm 324.5$
300	$4641.6 \pm 49.8$	$2346.6 \pm 51.1$	$117.9 \pm 1.5$	2455	$2464.50 \pm 51.11$
350	$2112.1 \pm 21.1$	$827.0 \pm 29.3$	$59.11 \pm 0.97$	908	$886.15 \pm 29.31$
400	$945.8 \pm 9.3$	$343.0 \pm 18.4$	$32.88 \pm 0.69$	413	$375.86 \pm 18.38$
450	$457.5 \pm 4.5$	$144.5 \pm 11.8$	$14.13 \pm 0.42$	192	$158.64 \pm 11.81$
500	$226.7 \pm 2.2$	$77.8 \pm 8.6$	$7.76 \pm 0.30$	83	$85.55 \pm 8.60$
550	$118.2 \pm 1.2$	$28.3 \pm 5.2$	$3.89 \pm 0.21$	44	$32.18 \pm 5.17$
600	$64.65 \pm 0.64$	$13.2 \pm 3.5$	$2.29 \pm 0.17$	28	$15.53 \pm 3.54$
650	$36.25 \pm 0.36$	$9.5 \pm 3.0$	$1.18 \pm 0.12$	18	$10.65 \pm 3.00$
700	$21.18 \pm 0.21$	$4.7 \pm 2.1$	$0.85 \pm 0.10$	6	$5.58 \pm 2.12$
750	$12.56 \pm 0.12$	$1.8 \pm 1.3$	$0.514 \pm 0.091$	4	$2.32 \pm 1.28$
800	$7.412 \pm 0.073$	$0.90 \pm 0.90$	$0.317 \pm 0.067$	3	$1.22 \pm 0.90$
850	$4.591 \pm 0.045$	$0.000^{+1.14}_{-0.00}$	$0.117 \pm 0.029$	2	$0.117^{+1.140}_{-0.029}$
900	$2.853 \pm 0.028$	$0.000^{+1.14}_{-0.00}$	$0.076 \pm 0.024$	1	$0.076^{+1.140}_{-0.024}$
950	$1.791 \pm 0.017$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$
1000	$1.272 \pm 0.011$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$
1050	$0.8788 \pm 0.0074$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$
1100	$0.6063 \pm 0.0049$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$
1150	$0.4196 \pm 0.0032$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$
1200	$0.2894 \pm 0.0021$	$0.000^{+1.14}_{-0.00}$	$0.069 \pm 0.023$	1	$0.069^{+1.140}_{-0.023}$

## 1054 L.6 Results at $evjj$ final selection

1055 The same final selection cuts described in Section 4.4.3 are applied to this version of  
 1056 the  $evjj$  analysis. The resulting background predictions at each level of final selection  
 1057 are given in Table 38.

1058 It is notable that this background estimate predicts  $10.65 \pm 3.00$  (stat) events at the fi-  
 1059 nal selection optimized for leptoquarks with a mass of 650 GeV. This is in agreement  
 1060 with  $7.54 \pm 1.20$  (stat) background events predicted by the main analysis. However,  
 1061 because the central value of the data-driven background prediction is higher, and be-  
 1062 cause the statistical uncertainty of the data-driven background prediction is larger,  
 1063 the significance of the excess at final selection for leptoquarks with a mass of 650  
 1064 GeV is less significant under the data-driven background prediction.

## 1065 L.7 Results and conclusions

1066 As shown in Sections L.2 and L.5, the method described in this appendix suc-  
 1067 cessfully models all of the important variables at preselection in both the  $eejj$  and  $evjj$  anal-  
 1068 yses.

1069 In the region of the excess (i.e. after applying the final selection optimized for a lep-  
 1070 toquark mass of 650 GeV), the purely data-driven background estimate agrees with  
 1071 the Monte Carlo background estimate within one standard deviation of statistical  
 1072 uncertainty. This result confirms that the excess is not due strictly to a deficiency in  
 1073 the Monte Carlo background estimate.

1074 It is possible to calculate limits using the results from Tables 36 and 38. The system-  
 1075 atic uncertainties used to calculate these limits include the uncertainties associated  
 1076 with the trigger efficiencies, the statistical uncertainties associated with the recon-  
 1077 struction efficiencies, the uncertainty associated with the QCD multijet background  
 1078 estimate, and the statistical uncertainty of the muon and QCD datasets (which dom-  
 1079 inates). These limits are shown in Figure 94.

1080 The significance in the excess at the final selection optimized for a leptoquark of  
 1081 mass 650 GeV is lower in both the  $eejj$  and the  $evjj$  analyses. This is for two reasons.

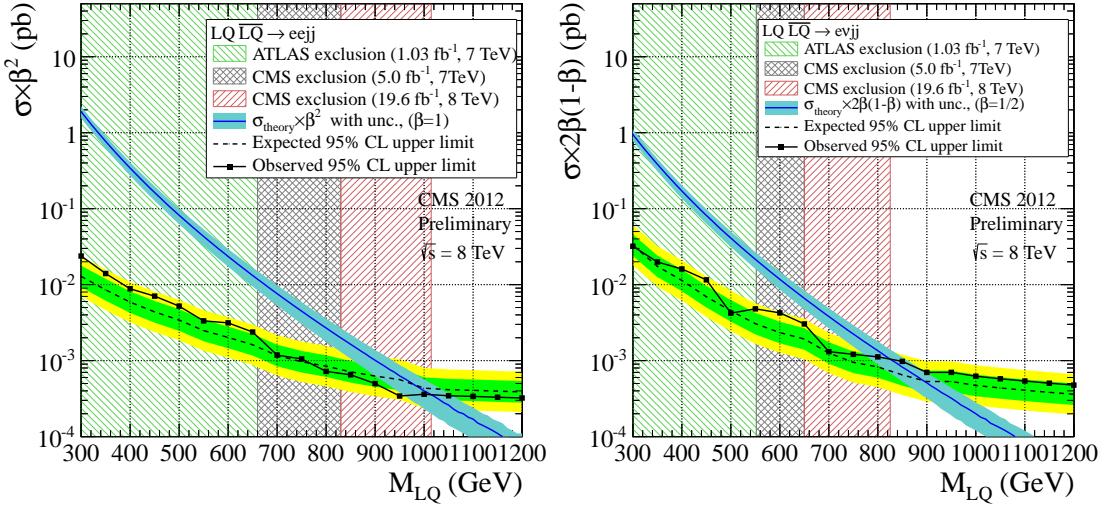


Figure 94: The expected and observed upper limit at 95% C.L. on the LQ pair production cross section times  $\beta^2$  in the top plot ( $2\beta(1 - \beta)$  in the bottom plot) as a function of the LQ mass obtained with the  $eejj$  ( $evjj$ ) analysis. In both plots, background predictions are obtained using purely data-driven methods. The dark blue curve and the light blue band represent, respectively, the theoretical LQ pair production cross section and the uncertainties due to the choice of PDF and renormalization/factorization scales.

First, while the results for the two estimates agree within one standard deviation of statistical uncertainty, the central value of this data-driven background estimate is higher than the central value of the Monte Carlo background estimate at the final selection optimized for a leptoquark of mass 650 GeV. Second, the statistical uncertainties associated with this data-driven background estimate are much larger than the statistical uncertainties associated with the Monte Carlo background estimate.

1082  
1083  
1084  
1085  
1086  
1087