Inclusive search for squarks and gluinos in pp collisions at s=7TeV

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Inclusive search for squarks and gluinos in \( pp \) collisions at \( \sqrt{s} = 7 \) TeV

S. Chatrchyan et al.*
(CMS Collaboration)
(Received 7 July 2011; published 11 January 2012)

A search is performed for heavy particle pairs produced in \( \sqrt{s} = 7 \) TeV proton-proton collisions with 35 pb\(^{-1}\) of data collected by the CMS experiment at the LHC. The search is sensitive to squarks and gluinos of generic supersymmetry models, provided they are kinematically accessible, with minimal assumptions on properties of the lightest superpartner particle. The kinematic consistency of the selected events is tested against the hypothesis of heavy particle pair production using the dimensionless \textit{razor} variable \( R \), related to the missing transverse energy \( E_T^{\text{miss}} \). The new physics signal is characterized by a broad peak in the distribution of \( M_R \), an event-by-event indicator of the heavy particle mass scale. This new approach is complementary to \( E_T^{\text{miss}} \)-based searches. After background modeling based on data, and background rejection based on \( R \) and \( M_R \), no significant excess of events is found beyond the standard model expectations. The results are interpreted in the context of the constrained minimal supersymmetric standard model as well as two simplified supersymmetry models.

DOI: 10.1103/PhysRevD.85.012004
PACS numbers: 13.75.Cs

I. INTRODUCTION

Models with softly broken supersymmetry (SUSY) [1–5] predict superpartners of the standard model (SM) particles. Experimental limits from the Tevatron and LEP showed that superpartner particles, if they exist, are significantly heavier than their SM counterparts. Proposed experimental searches for \( R \)-parity conserving SUSY at the Large Hadron Collider (LHC) have therefore focused on a combination of two SUSY signatures: multiple energetic jets and/or leptons from the decays of pair-produced squarks and gluinos, and large missing transverse energy (\( E_T^{\text{miss}} \)) from the two weakly interacting lightest superpartners (LSP) produced in separate decay chains.

In this paper, a new approach is presented that is inclusive not only for SUSY but also in the larger context of physics beyond the standard model. The focal point for this novel \textit{razor} analysis [6] is the production of pairs of heavy particles (of which squarks and gluinos are examples), whose masses are significantly larger than those of any SM particle. The analysis is designed to kinematically discriminate the pair production of heavy particles from SM backgrounds, without making strong assumptions about the \( E_T^{\text{miss}} \) spectrum or details of the decay chains of these particles. The baseline selection requires two or more reconstructed objects, which can be calorimetric jets, isolated electrons, or isolated muons. These objects are grouped into two \textit{megajets}. The \textit{razor} analysis tests the consistency, event by event, of the hypothesis that the two megajets represent the visible portion of the decays of two heavy particles. This strategy is complementary to traditional searches for signals in the tails of the \( E_T^{\text{miss}} \) distribution [7–16] and is applied to data collected with the Compact Muon Solenoid (CMS) detector from \( pp \) collisions at \( \sqrt{s} = 7 \) TeV corresponding to an integrated luminosity of 35 pb\(^{-1}\).

II. THE CMS APPARATUS

A description of the CMS detector can be found elsewhere [17]. A characteristic feature of the CMS detector is its superconducting solenoid magnet, of 6 m internal diameter, providing a field of 3.8 T. The silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass/scintillator hadron calorimeter (HCAL) are contained within the solenoid. Muons are detected in gas-ionization chambers embedded in the steel return yoke. The ECAL has an energy resolution of better than 0.5% above 100 GeV. The HCAL combined with the ECAL measures the jet energy with a resolution \( \Delta E/E = 100%/\sqrt{E/\text{GeV}} \oplus 5\% \).

CMS uses a coordinate system with the origin located at the nominal collision point, the \( x \) axis pointing towards the center of the LHC, the \( y \) axis pointing up (perpendicular to the LHC plane), and the \( z \) axis along the counterclockwise beam direction. The azimuthal angle \( \phi \) is measured with respect to the \( x \) axis in the \( xy \) plane and the polar angle \( \theta \) is defined with respect to the \( z \) axis. The pseudorapidity is \( \eta = -\ln[\tan(\theta/2)] \).

III. THE RAZOR ANALYSIS

The pair production of two heavy particles, each decaying to an unseen LSP plus jets, gives rise to a generic SUSY-like signal. Events in this analysis are forced into a dijet topology by combining all jets in the event into two...
Lorentz boost factor is defined by momentum from initial-state radiation. The longitudinal at threshold and if the CM system had no overall transverse CM frame for signal events, if the squarks were produced two megajets’ three-momenta. The longitudinally boosted frame that equalizes the magnitude of the momenta, respectively. The relation between squark pair production is illustrated in Fig. 1. To the extent that the signal is made assuming the dijet signal topology, replacing frame. Instead, an approximate event-by-event reconstruction is possible to reconstruct the actual CM frame. The R frame would be the CM frame for signal events, if the squarks were produced at threshold and if the CM system had no overall transverse momentum from initial-state radiation. The longitudinal Lorentz boost factor is defined by

\[ \beta_R = \frac{E^j - E^j}{p_T^j - p_T^j}, \]

where \( E^j, E^j \) and \( p_T^j, p_T^j \) are the megajet energies and longitudinal momenta, respectively. The relation between the laboratory frame, R frame, and CM frame view of squark pair production is illustrated in Fig. 1. To the extent that the R frame matches the true CM frame, the maximum value of the scalar sum of the megajets’ transverse momenta \( p_T^j, p_T^j \) is \( M_\Delta \) for signal events. The maximum value of the \( E_T^{\text{miss}} \) is also \( M_\Delta \). A transverse mass \( M_R^T \) is defined whose maximum value for signal events is also \( M_\Delta \) in the limit that the R and CM frames coincide:

\[ M_R^T = \frac{E_T^{\text{miss}}(p_T^j + p_T^j) - E_T^{\text{miss}} \cdot (\tilde{p}_T^1 + \tilde{p}_T^2)}{2}. \]

The event-by-event estimator of \( M_\Delta \) is

\[ M_R = 2|\tilde{p}_T^1| = 2|\tilde{p}_T^2|, \]

where \( \tilde{p}_T^1 \) and \( \tilde{p}_T^2 \) are the three-momenta of the megajets in the R frame. For signal events in the limit where the R frame and the true CM frame coincide, \( M_R \) equals \( M_\Delta \), and more generally \( M_R \) is expected to peak around \( M_\Delta \) for signal events. For QCD dijet and multijet events, the only
relevant scale is $\sqrt{s}$, the CM energy of the partonic subprocess.

The search for an excess of signal events in a tail of a distribution is thus recast as a search for a peak on top of a steeply falling SM residual tail in the $M_R$ distribution. To extract the peaking signal, the QCD multijet background needs to be reduced to manageable levels. This is achieved using the razor variable $R$ defined as

$$R = \frac{M_R^b}{M_R^T}$$

(6)

Since for signal events $M_R^b$ has a maximum value of $M_\Delta$ (i.e., a kinematic edge), $R$ has a maximum value of approximately 1 and the distribution of $R$ for signal events peaks around 0.5. These properties motivate the appropriate kinematic requirements for the signal selection and background reduction. It is noted that, while $M_R^b$ and $M_R^T$ measure the same scale (one as an end point, the other as a peak), they are largely uncorrelated for signal events, as shown in Fig. 2. In this figure, the $W$ + jets and $t\bar{t}$ + jets backgrounds peak at $M_R$ values partially determined by the $W$ and top quark masses, respectively. The fact that the figure exhibits the expected qualitative behavior for both signal and background shows that the megajet reconstruction, though approximate, yields a reasonable description of the underlying kinematics in simulated processes.

In this analysis the SM background shapes and normalizations are obtained from data. The backgrounds are extracted from control regions in the $R$ and $M_R$ distributions dominated by SM processes. Initial estimates of the background distributions in these regions are obtained from the individual simulated background components, but their shapes and normalizations are then corrected using data. As will be demonstrated in Sec. V, not only are the SM backgrounds greatly reduced by a threshold requirement on $R$, but also the $M_R$ distributions of the residual backgrounds have simple exponential shapes. The simplicity of the background modeling, together with the peaking property of potential signals, are both in contrast to more traditional searches for SUSY based on high-$p_T$ and $E_T^{\text{miss}}$ tails.

The analysis flow is as follows:

1. The inclusive data sets are collected using the electron, muon, and hadronic-jet triggers.

---

**FIG. 2** (color online). Scatter plot in the ($M_R$, $R$) plane for simulated events: (top left) QCD multijet, (top right) $W$ + jets, (bottom left) $t\bar{t}$ + jets, and (bottom right) the SUSY benchmark model LM1 [43] with $M_\Delta = 597$ GeV. The yields are normalized to an integrated luminosity of 35 pb$^{-1}$. The bin size is (20 GeV $\times$ 0.015).
These data sets are examined for the presence of a well-identified isolated electron or muon, irrespective of the trigger path. Based on the presence or absence of such a lepton, each event is assigned to one of three disjoint event samples, referred to as the electron, muon, and hadronic boxes. These boxes serve as controls of processes in the SM with leptons, jets, and neutrinos, e.g. QCD multijet, $W + \text{jets}$ or $Z + \text{jets}$, and $t + X$. The diboson background is found to be negligible. Exclusive multilepton boxes are also defined but are not sufficiently populated to be used in this analysis.

Megajets are constructed for events passing a baseline kinematic selection, and the $R$ and $M_R$ event variables are computed. In the electron box, electrons are clustered with jets in the definition of the megajets. Jets matched to these electrons are removed to avoid double counting. In the muon box, muons are included in the megajet clustering. In order to characterize the distribution of the SM background events in the $(M_R, R)$ plane, a kinematic region is identified in the lepton boxes that is dominated by $W(\ell \nu) + \text{jets}$. Another region is found that is dominated by the sum of the non-QCD backgrounds. Events remaining in the hadronic box primarily consist of QCD multijet, $Z(\ell \ell') + \text{jets}$, $W(\ell \nu) + \text{jets}$, and $t + X$ events that produce $\ell + \text{jets} + E_T^{\text{miss}}$ final states with charged leptons that do not satisfy the electron or muon selections. The shapes and normalizations of these non-QCD background processes in the hadronic box are estimated using the results from the lepton boxes in appropriate regions in the $(M_R, R)$ plane.

In addition to control regions defined in the boxes, three additional control samples are used to estimate the QCD background. The QCD background shape and normalization in each of the lepton boxes is extracted by reversing the lepton isolation requirements to obtain control samples dominated by QCD background. The QCD background in the hadronic box is estimated using a QCD control sample collected with prescaled jet triggers. The large-$R$ and high-$M_R$ regions of all boxes are signal candidate regions not used for the background estimates. Above a given $R$ threshold, the $M_R$ distribution of the backgrounds observed in the data is well modeled by simple exponential functions. Having determined the $R$ and $M_R$ shape and normalization of the backgrounds in the control regions, the SM yields are extrapolated to the large-$R$ and high-$M_R$ signal candidate regions for each box.

IV. EVENT SELECTION

The analysis uses data sets recorded with triggers based on the presence of an electron, a muon, or on $H_T$, the uncorrected scalar sum of the transverse energy of jets reconstructed at the trigger level. Prescaled jet triggers with low thresholds are used for the QCD multijet background estimation in the hadronic box.

The analysis is guided by studies of Monte Carlo (MC) event samples generated with the PYTHIA [18] and MADGRAPH [19] programs, simulated using the CMS GEANT-based [20] detector simulation, and then processed by the same software used to reconstruct real collision data. Events with QCD multijet, top quarks, and electroweak bosons were generated with MADGRAPH interfaced with PYTHIA for parton showering, hadronization, and

![FIG. 3 (color online).](image)

(left) $M_R$ distributions for different values of the $R$ threshold for data events in the QCD multijet control sample. Fits of the $M_R$ distribution to an exponential function and an asymmetric Gaussian at low $M_R$, are shown as dotted black curves. (right) The exponential slope $S$ from fits to the $M_R$ distribution, as a function of the square of the $R$ threshold for data events in the QCD control sample.
underlying event description. To generate Monte Carlo samples for SUSY, the mass spectrum was first calculated with SOFTSUSY [21] and the decays with SUSYHIT [22]. The PYTHIA program was used with the SLHA interface [23] to generate the events. The generator level cross section and the $K$ factors for the next-to-leading order (NLO) cross section calculation were computed using PROSPINO [24].

Events are required to have at least one good reconstructed interaction vertex [25]. When multiple vertices are found, the one with the highest associated $p_T$ is used. Jets are reconstructed offline from calorimeter energy deposits using the infrared-safe anti-$k_T$ [26] algorithm with radius parameter 0.5. Jets are corrected for the nonuniformity of the calorimeter response in energy and $\eta$ using corrections derived with the simulation and are required to have $p_T > 30$ GeV and $|\eta| < 3.0$. The jet energy scale uncertainty for these corrected jets is 5% [27]. The $E_T^{\text{miss}}$ is reconstructed using the particle flow algorithm [28].

The electron and muon reconstruction and identification criteria are described in [29]. Isolated electrons and muons are required to have $p_T > 20$ GeV and $|\eta| < 2.5$ and 2.1, respectively, and to satisfy the selection requirements from [29]. The typical lepton trigger and reconstruction efficiencies are 98% and 99%, respectively, for electrons and 95% and 98% for muons.

The reconstructed hadronic jets, isolated electrons, and isolated muons are grouped into two megajets, when at least two such objects are present in the event. The megajets are constructed as a sum of the four-momenta of their constituent objects. After considering all possible partitions of the objects into two megajets, the combination minimizing the invariant masses summed in quadrature of the resulting megajets is selected among all combinations for which the $R$ frame is well defined.

After the construction of the two megajets, the boost variable $S = a + bR^2$ is computed; due to the approximations mentioned above, $|\beta_R|$ can fall in an unphysical region ($\geq 1$) for signal or background events; these events are removed. The additional requirement $|\beta_R| \leq 0.99$ is imposed to remove events for which the razor variables become singular. This

FIG. 4 (color online). (left) $M_R$ distributions for different values of the $R$ threshold from data events selected in the MU (upper) and ELE (lower) boxes. Dotted curves show the results of fits using two independent exponential functions and an asymmetric Gaussian at low $M_R$. (right) The slope $S$ of the first exponential component as a function of the square of the $R$ threshold in the MU (upper) and ELE (lower) boxes. The dotted lines show the results of the fits to the form $S = a + bR^2$. 

012004-5
requirement is typically 85% efficient for simulated SUSY events, and 80 to 85% efficient for background events that otherwise pass the baseline selection. The azimuthal angular difference between the megajets is required to be less than 2.8 radians; this requirement suppresses nearly back-to-back QCD dijet events. These requirements define the inclusive baseline selection. After this selection, the signal efficiency in the constrained minimal supersymmetric standard model (CMSSM) [30–33] parameter space for a gluino mass of \( \sim 600 \) GeV is over 50%.

V. BACKGROUND ESTIMATION

In traditional searches for SUSY, QCD multijet production is an especially daunting background because of its very high cross section and complicated modeling of its high-\( p_T \) and \( E_{\text{T}}^{\text{miss}} \) tails, including contributions from spurious instrumental effects. In this analysis a cut on \( R \) makes it possible to model the QCD multijet background and confine its contribution to the low-\( M_R \) region.

Apart from QCD multijet backgrounds, the remaining backgrounds in the lepton and hadronic boxes are processes with genuine \( E_{\text{T}}^{\text{miss}} \) due to energetic neutrinos and leptons from massive vector boson decays (including \( W \) bosons from top quark decays). After applying an \( R \) threshold, the \( M_R \) distributions in the lepton and hadronic boxes are very similar for these backgrounds; this similarity is exploited in the modeling and normalization of these backgrounds.

A. QCD multijet background

The QCD multijet control sample for the hadronic box is defined from event samples recorded with prescaled jet triggers and passing the baseline analysis selection for events without a well-identified isolated electron or muon. The trigger requires at least two jets with an average uncorrected \( p_T > 15 \) GeV. Because of the low jet threshold, the QCD multijet background dominates this sample for low \( M_R \), thus allowing the extraction of the \( M_R \) shapes with different \( R \) thresholds for QCD multijet events. These shapes are corrected for the \( H_T \) trigger turn-on efficiency.

The \( M_R \) distributions for events in the QCD control sample selection, for different values of the \( R \) threshold, are shown in Fig. 3 (left). The \( M_R \) distribution is exponentially falling, after a turn-on at low \( M_R \) resulting from the \( p_T \) threshold requirement on the jets entering the megajet calculation. After the turn-on which is fitted with an asymmetric Gaussian, the exponential region of these distributions is fitted for each value of \( R \) to extract the exponential slope, denoted by \( S \). The value of \( S \) that maximizes the likelihood in the exponential fit is found to be a linear function of \( R^2 \), as shown in Fig. 3 (right); fitting \( S \) to the form \( S = a + bR^2 \) determines the values of the shape parameters \( a \) and \( b \).

When measuring the exponential slopes of the \( M_R \) distributions as a function of the \( R \) threshold, the correlations due to events satisfying multiple \( R \) threshold requirements are neglected. The effect of these correlations on the measurement of the slopes is studied by using pseudoexperiments and is found to be negligible.

To measure the shape of the QCD background component in the lepton boxes, the corresponding lepton trigger data sets are used with the baseline selection and reversed lepton isolation criteria. The QCD background component in the lepton boxes is found to be negligible.

The \( R \) threshold shapes the \( M_R \) distribution in a simple therefore predictable way: for higher \( M_R \) the \( M_R \) distribution is a simple exponential, fully described in terms of the value of the \( R \) threshold and the two shape parameters \( a \) and \( b \). Event selections with combined \( R \) and \( M_R \) thresholds are found to suppress jet mismeasurements, including severe mismeasurements of the electromagnetic or hadronic component of the jet energy, or other anomalous calorimetric noise signals such as the ones described in [34,35].

FIG. 5 (color online). The \( M_R \) distributions with \( R > 0.45 \) in the ELE (left) and MU (right) boxes for data (points) and backgrounds (curves). The bands show the uncertainties of the background predictions.
TABLE I. The number of predicted background events in the ELE and MU boxes for \( R > 0.45 \) and \( M_R > 500 \text{ GeV} \) and the number of events observed in data.

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELE box</td>
<td>0.63 ± 0.23</td>
<td>0</td>
</tr>
<tr>
<td>MU box</td>
<td>0.51 ± 0.20</td>
<td>3</td>
</tr>
</tbody>
</table>

B. \( W + \) jets, \( Z + \) jets, and \( t + X \) backgrounds

Using the muon (MU) and electron (ELE) boxes defined in Sec. III, \( M_R \) intervals dominated by \( W(\ell \nu) + \) jets events are identified for different \( R \) thresholds. In both simulated and data events, the \( M_R \) distribution is well described by two independent exponential components. The first component of \( W(\ell \nu) + \) jets corresponds to events where the highest \( p_T \) object in one of the megajets is the isolated electron or muon; the second component consists of events where the leading object in both megajets is a jet, as is typical also for the \( t + X \) background events. The first component of \( W(\ell \nu) + \) jets can be measured directly in data, because it dominates over all other backgrounds in a control region of lower \( M_R \) set by the \( R \) threshold. At higher values of \( M_R \), the first component of \( W(\ell \nu) + \) jets falls off rapidly; in simulated events this is seen to be a result of the anticorrelation between the neutrino \( p_T \) (necessary to pass the \( R \) threshold) and the lepton \( p_T \) (necessary to provide the second megajet), for a given value of the \( W \) \( p_T \) (necessary to achieve a given value of \( M_R \)). The remaining background is instead dominated by the sum of \( t + X \) and the second component of \( W(\ell \nu) + \) jets; this defines a second control region of intermediate \( M_R \) set by the \( R \) threshold.

Using these two control regions in a given box, a simultaneous fit determines both exponential slopes along with the absolute normalization of the first component of \( W(\ell \nu) + \) jets and the relative normalization of the sum of the second component of \( W(\ell \nu) + \) jets with the other backgrounds. The \( M_R \) distributions as a function of \( R \) are shown in Fig. 4 (left). The slope parameters characterizing the exponential behavior of the first \( W(\ell \nu) + \) jets component are shown in Fig. 4 (right); they are consistent within uncertainties between the electron and muon channels. The values of the parameters \( a \) and \( b \) that describe the \( R^2 \) dependence of the slope are in good agreement with the values extracted from simulated \( W(\ell \nu) + \) jets events.

Given the fits just described to the first \( W(\ell \nu) + \) jets component, we define data/MC ratios \( \rho(a)_1^{\text{data/MC}} \), \( \rho(b)_1^{\text{data/MC}} \) of the first component slope parameters \( a \), \( b \) measured in the MU and ELE boxes. These \( \rho_1^{\text{data/MC}} \) are then used as correction factors for the shape parameters \( a \) and \( b \) of the \( Z + \) jets and \( t + X \) backgrounds as extracted from simulated samples for the MU and ELE boxes. Because of the agreement found in data between the electron and muon channels, the values of \( \rho(a)_1^{\text{data/MC}} \), \( \rho(b)_1^{\text{data/MC}} \) as measured in the leptonic boxes are combined, yielding

\[
\rho(a)_1^{\text{data/MC}} = 0.97 \pm 0.02; \\
\rho(b)_1^{\text{data/MC}} = 0.97 \pm 0.02,
\]

where the quoted uncertainties are determined from the fits.

The data/MC correction factors for the second component of \( W(\ell \nu) + \) jets in the MU and ELE boxes are measured in these boxes using a lepton-as-neutrino treatment of leptonic events. Here the electron or muon is excluded from the megajet reconstruction, kinematically mimicking the presence of an additional neutrino. With the lepton-as-neutrino treatment in the MU and ELE boxes, only one exponential component is observed both in data and in \( W(\ell \nu) + \) jets simulated events. In the simulation, the value of this single exponential component slope is found to agree with the value for the second component of \( W(\ell \nu) + \) jets obtained in the default treatment.

The combined data/MC correction factors measured using this lepton-as-neutrino treatment are

\[
\rho(a)_2^{\text{data/MC}} = 1.01 \pm 0.02; \\
\rho(b)_2^{\text{data/MC}} = 0.94 \pm 0.07.
\]

TABLE II. Summary of the uncertainties on the background predictions for the ELE and MU boxes and their relative magnitudes. The range in the Monte Carlo uncertainties is owing to the different statistical precisions of the simulated background samples.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Relative magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope parameter ( a )</td>
<td>Systematic bias from correlations in fits</td>
<td>5%</td>
</tr>
<tr>
<td>Slope parameter ( b )</td>
<td>Systematic bias from correlations in fits</td>
<td>10%</td>
</tr>
<tr>
<td>Slope parameter ( a )</td>
<td>Uncertainty from Monte Carlo</td>
<td>1%–10%</td>
</tr>
<tr>
<td>Slope parameter ( b )</td>
<td>Uncertainty from Monte Carlo</td>
<td>1%–10%</td>
</tr>
<tr>
<td>( \rho(a)_1^{\text{data/MC}} )</td>
<td>Data fit</td>
<td>3%</td>
</tr>
<tr>
<td>( \rho(b)_1^{\text{data/MC}} )</td>
<td>Data fit</td>
<td>3%</td>
</tr>
<tr>
<td>Normalization</td>
<td>Systematic + statistical component</td>
<td>3%–8%</td>
</tr>
<tr>
<td>( f_W )</td>
<td>Extracted from fit (( W ) only)</td>
<td>30%</td>
</tr>
<tr>
<td>( W/\ell) cross section ratio</td>
<td>CMS measurements (top only)</td>
<td>40%</td>
</tr>
<tr>
<td>( W/Z ) cross section ratio</td>
<td>CMS measurements (( Z ) only)</td>
<td>19%</td>
</tr>
</tbody>
</table>
Background corrections for the hadronic box

The ratios \( \rho_{1,2}^{\text{data/MC}} \) obtained from the leptonic boxes are taken as correction factors for the shape of the first component of \( W(\ell \nu) + \text{jets} \) as extracted from simulated samples for the hadronic box. The data/MC correction factors \( \rho_{2}^{\text{data/MC}} \) are used for the second component of \( W(\ell \nu) + \text{jets} \) in the HAD box, as well as the \( Z(\nu \bar{\nu}) + \text{jets} \) and \( t + X \) backgrounds in the HAD box. For the final background prediction the magnitude of the relative normalization between the two \( W(\ell \nu) + \text{jets} \) components, denoted \( f^W \), is determined from a binned maximum likelihood fit in the region \( 200 < M_R < 400 \) GeV.

VI. RESULTS

A. Lepton box background predictions

Having extracted the \( M_R \) shape of the \( W + \text{jets} \) and \( Z + \text{jets} \) backgrounds, their relative normalization is set from the \( W + Z \) cross sections measured by CMS in electron and muon final states [29]. Similarly, the normalization of the \( t\bar{t} \) background relative to \( W + \text{jets} \) is taken from the \( t\bar{t} \) cross section measured by CMS in the dilepton channel [36]. The measured values of these cross sections are summarized below:

\[
\begin{align*}
\sigma(pp \rightarrow WX) \times B(W \rightarrow \ell \nu) & = 9.951 \pm 0.073(\text{stat.}) \pm 0.280(\text{syst.}) \pm 1.095(\text{lumi}) \text{ nb}, \\
\sigma(pp \rightarrow ZX) \times B(Z \rightarrow \ell \ell) & = 0.931 \pm 0.026(\text{stat.}) \pm 0.023(\text{syst.}) \pm 0.102(\text{lumi}) \text{ nb}, \\
\sigma(pp \rightarrow t\bar{t}) & = 194 \pm 72(\text{stat.}) \pm 24(\text{syst.}) \pm 21(\text{lumi}) \text{ pb}.
\end{align*}
\] (9)

For an \( R > 0.45 \) threshold the QCD background is virtually eliminated. The region \( 125 < M_R < 175 \) GeV where the QCD contribution is negligible and the \( W(\ell \nu) + \text{jets} \) component is dominant is used to fix the overall normalization of the total background prediction. The final background prediction in the ELE and MU boxes for \( R > 0.45 \) is shown in Fig. 5.

The number of events with \( M_R > 500 \) GeV observed in data and the corresponding number of predicted background events are given in Table I for the ELE and MU boxes. Agreement between the predicted and observed yields is found. The \( p \) value of the measurement in the MU box is 0.1, given the predicted background (with its statistical and systematic uncertainties) and the observed number of events. A summary of the uncertainties entering the background measurements is presented in Table II.

B. Hadronic box background predictions

The procedure for estimating the total background predictions in the hadronic box is summarized as follows:

(i) Construct the non-QCD background shapes in \( M_R \) using measured values of \( a \) and \( b \) from simulated events, applying correction factors derived from data control samples, and taking into account the \( H_T \) trigger turn-on efficiency.

(ii) Set the relative normalizations of the \( W + \text{jets} \), \( Z + \text{jets} \), and \( t + X \) backgrounds using the relevant inclusive cross section measurements from CMS [Eq. (9)].

(iii) Set the overall normalization by measuring the event yields in the lepton boxes, corrected for

\[
\begin{array}{c|c|c}
M_R & \text{Predicted} & \text{Observed} \\
\hline
M_R > 500 \text{ GeV} & 5.5 \pm 1.4 & 7 \\
\end{array}
\]

TABLE III. Predicted and observed yields for \( M_R > 500 \) GeV with \( R > 0.5 \) in the HAD box.
TABLE IV. Summary of uncertainties entering the background predictions for the HAD box.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Relative magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope parameter $a$</td>
<td>Systematic bias from correlations in fits</td>
<td>5%</td>
</tr>
<tr>
<td>Slope parameter $b$</td>
<td>Systematic bias from correlations in fits</td>
<td>10%</td>
</tr>
<tr>
<td>Slope parameter $a$</td>
<td>Uncertainty from Monte Carlo</td>
<td>1%–10%</td>
</tr>
<tr>
<td>Slope parameter $b$</td>
<td>Uncertainty from Monte Carlo</td>
<td>1%–10%</td>
</tr>
<tr>
<td>$\rho(a)_{\text{data/MC}}$</td>
<td>Data fit</td>
<td>3%</td>
</tr>
<tr>
<td>$\rho(b)_{\text{data/MC}}$</td>
<td>Data fit</td>
<td>3%</td>
</tr>
<tr>
<td>Normalization</td>
<td>Systematic + statistical component</td>
<td>8%</td>
</tr>
<tr>
<td>Trigger parameters</td>
<td>Systematic from fit pseudoexperiments</td>
<td>2%</td>
</tr>
<tr>
<td>$f_W$</td>
<td>Extracted from fit (W only)</td>
<td>13%</td>
</tr>
<tr>
<td>$W/t\bar{t}$ cross section ratio</td>
<td>CMS measurements (top only)</td>
<td>40%</td>
</tr>
<tr>
<td>$W/Z$ cross section ratio</td>
<td>CMS measurements (Z only)</td>
<td>19%</td>
</tr>
</tbody>
</table>

lepton reconstruction and identification efficiencies. The shapes and normalizations of all the non-QCD backgrounds are now fixed.

(iv) The shape of the QCD background is extracted, as described in Sec. VA, and its normalization in the HAD box is determined from a fit to the low-$M_R$ region, as described below.

The final hadronic box background prediction is calculated from a binned likelihood fit of the total background shape to the data in the interval $80 < M_R < 400$ GeV with all background normalizations and shapes fixed, except for the following free parameters: (i) the $H_T$ trigger turn-on shapes, (ii) $f_W$ as introduced in Sec. V B, and (iii) the overall normalization of the QCD background. A set of pseudoexperiments is used to test the overall fit for coverage of the various floated parameters and for systematic biases. A 2% systematic uncertainty is assigned to the high-$M_R$ background prediction that encapsulates systematic effects related to the fitting procedure. Figure 6 shows the final hadronic box background prediction with all uncertainties on this prediction included for $R > 0.5$. The observed $M_R$ distribution is consistent with the predicted one over the entire $M_R$ range. The predicted and observed background yields in the high-$M_R$ region are summarized in Table III. A summary of the uncertainties entering these background predictions is listed in Table IV. A larger $R$ requirement is used in the HAD box analysis due to the larger background.

VII. LIMITS IN THE CMSSM PARAMETER SPACE

Having observed no significant excess of events beyond the SM expectations, we extract a model-independent 95% confidence level (C.L.) limit on the number of signal events. This limit is then interpreted in the parameter spaces of SUSY models.

The likelihood for the number of observed events $n$ is modeled as a Poisson function, given the sum of the number of signal events ($s$) and the number of background events. A posterior probability density function $P(s)$ for the signal yield is derived using Bayes theorem, assuming a flat prior for the signal and a log-normal prior for the background.

The model-independent upper limit is derived by integrating the posterior probability density function between 0 and $s^*$ so that $\int_0^{s^*} P(s)\,ds = 0.95$. The observed upper limit in the hadronic box is $s^* = 8.4$ (expected limit 7.2 ± 2.7); in the muon box $s^* = 6.3$ (expected limit 3.5 ± 1.1); and in the electron box $s^* = 2.9$ (expected limit 3.6 ± 1.1). For 10% of the pseudoexperiments in the muon box, the expected limit is higher than the observed. The stability of the result was studied with different choices of the signal prior. In particular, using the reference priors derived with the methods described in Ref. [37], the observed upper limits in the hadronic, muon, and electron boxes are 8.0, 5.3, and 2.9, respectively.

The results can be interpreted in the context of the CMSSM, which is a truncation of the full SUSY parameter space motivated by the minimal supergravity framework for spontaneous soft breaking of supersymmetry. In the CMSSM, the soft breaking parameters are reduced to five: three mass parameters $m_0$, $m_1/2$, and $A_0$ being, respectively, 5.3, 2.9, and 2.9.

TABLE V. Summary of the systematic uncertainties on the signal yield and totals for each of the event boxes. For the CMSSM scan the NLO signal cross section uncertainty is included.

<table>
<thead>
<tr>
<th>Box</th>
<th>MU</th>
<th>ELE</th>
<th>HAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JES</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Data/MC $\epsilon$</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>$\mathcal{L}$ [39]</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>Theory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISR</td>
<td>1%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>PDF</td>
<td>3%–6%</td>
<td>3%–6%</td>
<td>3%–6%</td>
</tr>
<tr>
<td>Subtotal</td>
<td>8%–9%</td>
<td>8%–9%</td>
<td>8%–9%</td>
</tr>
<tr>
<td>CMSSM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLO</td>
<td>16%–18%</td>
<td>16%–18%</td>
<td>16%–18%</td>
</tr>
<tr>
<td>Total</td>
<td>17%–19%</td>
<td>17%–19%</td>
<td>17%–19%</td>
</tr>
</tbody>
</table>
a universal scalar mass, a universal gaugino mass, and a universal trilinear scalar coupling, as well as \( \tan \beta \), the ratio of the up-type and down-type Higgs vacuum expectation values, and the sign of the supersymmetric Higgs mass parameter \( \mu \). Scanning over these parameters yields models which, while not entirely representative of the complete SUSY parameter space, vary widely in their superpartner spectra and thus in the dominant production channels and decay chains. This has the advantage that the razor limits address simultaneously a significant fraction of the possible

FIG. 7 (color online). Observed (solid curve) and expected (dot-dashed curve) 95% C.L. limits in the \((m_0, m_{1/2})\) CMSSM plane with \( \tan \beta = 3 \) (left column) and \( \tan \beta = 10 \) (right column), \( A_0 = 0 \), \( \text{sgn}(\mu) = +1 \) from the ELE/MU/HAD box selection (top to bottom) \((R > 0.45 \text{ for ELE/MU}, 0.5 \text{ for HAD}, M_0 \geq 500 \text{ GeV})\). The ± 1 standard deviation equivalent variations in the uncertainties are shown as a band around the expected limits.
as shown in Table V. The JES, ISR, and PDF uncertainties are relatively small owing to the insensitivity of the signal kinematics and final states. The signal efficiency in all three signal regions varies widely, from ~1 to 15%, depending on the SUSY model.

The upper limits are projected onto the \((m_0, m_{1/2})\) plane by comparing them with the predicted yields, and excluding any model if \(s(m_0, m_{1/2}) > s^*\). The systematic uncertainty on the signal yield (coming from the uncertainty on the luminosity, the selection efficiency, and the theoretical uncertainty associated with the cross section calculation) is modeled according to a log-normal prior. The uncertainty on the selection efficiency includes the effect of jet energy scale (JES) corrections, parton distribution function (PDF) uncertainties [38], and the description of initial-state radiation (ISR). All the effects are summed in quadrature as shown in Table V. The JES, ISR, and PDF uncertainties are relatively small owing to the insensitivity of the signal \(R\) and \(M_R\) distributions to these effects.

The observed limits from the ELE, MU, and HAD boxes are shown in Fig. 7, in the CMSSM \((m_0, m_{1/2})\) plane for the values \(\tan\beta = 3\) or \(\tan\beta = 10\), \(A_0 = 0\), \(\text{sgn}(\mu) = +1\), together with the 68% probability band for the expected limits, obtained by applying the same procedure to an ensemble of background-only pseudoexperiments. The band is computed around the median of the limit distribution. Observed limits are also shown in Fig. 8 for the values \(\tan\beta = 50\), \(A_0 = 0\), \(\text{sgn}(\mu) = +1\).

Figure 9 shows the same result in terms of 95% C.L. upper limits on the cross section as a function of the physical masses for two benchmark simplified models [13,40–42]: four-flavor squark pair production and gluino pair production. In the former, each squark decays to one quark and the LSP, resulting in final states with two jets and missing transverse energy, while in the latter each gluino decays directly to two light quarks and the LSP, giving events with four jets and missing transverse energy.

![Graphs showing limits on signal cross section](image-url)
VIII. SUMMARY

We performed a search for squarks and gluinos using a data sample of 35 pb$^{-1}$ integrated luminosity from $pp$ collisions at $\sqrt{s} = 7$ TeV, recorded by the CMS detector at the LHC. The kinematic consistency of the selected events was tested against the hypothesis of heavy particle pair production using the dimensionless razor variable $R$ related to the missing transverse energy $E_T^{miss}$, and $M_R$, an event-by-event indicator of the heavy particle mass scale. We used events with large $R$ and high $M_R$ in inclusive topologies.

The search relied on predictions of the SM backgrounds determined from data samples dominated by SM processes. No significant excess over the background expectations was observed, and model-independent upper limits on the numbers of signal events were calculated. The results were presented in the $(m_0, m_{1/2})$ CMSSM parameter space. For simplified models the results were given as limits on the production cross sections as a function of the squark, gluino, and LSP masses.

These results demonstrate the strengths of the razor analysis approach; the simple exponential behavior of the various SM backgrounds when described in terms of the razor variables is useful in suppressing these backgrounds and in making reliable estimates from data of the background residuals in the signal regions. Hence, the razor method provides an additional powerful probe in searching for physics beyond the SM at the LHC.

ACKNOWLEDGMENTS

We wish to congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC machine. We thank the technical and administrative staff at CERN and other CMS institutes. This work was supported by the Austrian Federal Ministry of Science and Research; the Belgian Fonds de la Recherche Scientifique, and Fonds voor Wetenschappelijk Onderzoek; the Brazilian Funding Agencies (CNPq, CAPES, FAPERJ, and FAPESP); the Bulgarian Ministry of Education and Science; CERN; the Chinese Academy of Sciences, the Ministry of Science and Technology, and National Natural Science Foundation of China; the Colombian Funding Agency (COLCIENCIAS); the Croatian Ministry of Science, Education and Culture, and Helsinki Institute of Physics; the Academy of Finland, Finnish Ministry of Education and Culture, and Helsinki Institute of Physics; the Institut National de Physique Nucléaire et de Physique des Particules/CNRS, and Commissariat à l’Énergie Atomique et aux Énergies Alternatives/CEA, France; the Bundesministerium für Bildung und Forschung, Deutsche Forschungsgemeinschaft, and Helmholtz-Gemeinschaft Deutscher Forschungszentren, Germany; the General Secretariat for Research and Technology, Greece; the National Scientific Research Foundation, and National Office for Research and Technology, Hungary; the Department of Atomic Energy and the Department of Science and Technology, India; the Institute for Studies in Theoretical Physics and Mathematics, Iran; the Science Foundation, Ireland; the Istituto Nazionale di Fisica Nucleare, Italy; the Korean Ministry of Education, Science and Technology and the World Class University program of NRF, Korea; the Lithuanian Academy of Sciences; the Mexican Funding Agencies (CINVESTAV, CONACYT, SEP, and UASLP-FAI); the Ministry of Science and Innovation, New Zealand; the Pakistan Atomic Energy Commission; the State Commission for Scientific Research, Poland; the Fundaçao para a Ciência e a Tecnologia, Portugal; JINR (Armenia, Belarus, Georgia, Ukraine, Uzbekistan); the Ministry of Science and Technologies of the Russian Federation, and Russian Ministry of Atomic Energy; the Ministry of Science and Technological Development of Serbia; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Swiss Funding Agencies (ETH Board, ETH Zurich, PSI, SNF, UniZH, Canton Zurich, and SER); the National Science Council, Taiwan; the Scientific and Technical Research Council of Turkey, and Turkish Atomic Energy Authority; the Science and Technology Facilities Council, UK; the U.S. Department of Energy, and the U.S. National Science Foundation. Individuals have received support from the Marie-Curie program and the European Research Council (European Union); the Leventis Foundation; the A.P. Sloan Foundation; the Alexander von Humboldt Foundation; the Associazione per lo Sviluppo Scientifico e Tecnologico del Piemonte (Italy); the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l’Industrie et dans l’Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); and the Council of Science and Industrial Research, India.


012004-12
INCLUSIVE SEARCH FOR SQUARKS AND GLUINOS IN . . .

INCLUSIVE SEARCH FOR SQUARKS AND GLUINOS IN...
aDeceased.
bAlso at CERN, European Organization for Nuclear Research, Geneva, Switzerland.
cAlso at Universidade Federal do ABC, Santo Andre, Brazil.
dAlso at Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France.
eAlso at Suez Canal University, Suez, Egypt.
fAlso at British University, Cairo, Egypt.
gAlso at Fayoum University, El-Fayoum, Egypt.
hAlso at Soltan Institute for Nuclear Studies, Warsaw, Poland.
iAlso at Massachusetts Institute of Technology, Cambridge, USA.
jAlso at Université de Haute-Alsace, Mulhouse, France.
kAlso at Brandenburg University of Technology, Cottbus, Germany.
lAlso at Moscow State University, Moscow, Russia.
mAlso at Institute of Nuclear Research ATOMKI, Debrecen, Hungary.

Also at Eötvös Loránd University, Budapest, Hungary.
oAlso at Tata Institute of Fundamental Research-HECR, Mumbai, India.
pAlso at University of Visva-Bharati, Santiniketan, India.
qAlso at Sharif University of Technology, Tehran, Iran.
rAlso at Isfahan University of Technology, Isfahan, Iran.
sAlso at Facoltà Ingegneria Università di Roma, Roma, Italy.