Exhaustive model selection in $b \to s\ell\ell$ decays: Pitting cross-validation against the Akaike information criterion

Srimoy Bhattacharya,[1,*] Aritra Biswas,[2,†] Soumitra Nandi,[2,‡] and Sunando Kumar Patra[2,§]

[1]*The Institute of Mathematical Sciences, C.I.T Campus, Taramani, Chennai 600 113, India*
[2]*Indian Institute of Technology, North Guwahati, Guwahati 781039, Assam, India*

In light of recent data, we study the new physics effects in the exclusive $b \to s\ell^+\ell^-$ decays from a model independent perspective. Different combinations of the dimension six effective operators along with their respective Wilson coefficients are chosen for the analysis. To find out the operator or sets of operators that can best explain the available data in this channel, we simultaneously apply popular model selection tools like cross-validation and the information theoretic approach like Akaike information criterion (AIC). There are one, two, and three-operator scenarios which survive the test and a left-handed quark current with vector muon coupling is common among them. This is also the only surviving one-operator scenario. Best-fit values and correlations of the new Wilson coefficients are supplied for all the selected scenarios. We find that the angular observables play the dominant role in the model selection procedure. We also note that while a left-handed quark current with axial-vector muon coupling is the only one-operator scenario able to explain the ratios $R_{K^{(*)}}$ (R_{K^*} for $q^2 \in [0.045, 1.1]$ GeV^2 in particular), there are also a couple of two operator scenarios that can simultaneously explain the measured $R_{K^{(*)}}$.

I. INTRODUCTION

Decays involving $b \to s\ell\ell$ transitions are suppressed in the standard model (SM). These decay modes are potentially sensitive to new physics effects. Whether the contributions appear at the tree or the loop level depends on the type of the new physics (NP). A lot of attention, both experimental and theoretical, have been given to $B \to K^{(*)}\mu^+\mu^-$ decays in the last couple of years. There are several angular observables associated with these decays, which are potentially sensitive to the NP effects and are measured by LHCb [1,2]. A couple of them have shown discrepancies with their respective SM predictions [3–6]. However, these angular observables are not free from hadronic uncertainties and it is fairly possible that the observed discrepancies are due to poorly known hadronic effects, e.g., see [7] for details. Furthermore, these modes offer theoretically clean observables like

$$R_H = \frac{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\Gamma(B \to H\mu^+\mu^-)}{dq^2}}{\int_{q_{\min}^2}^{q_{\max}^2} \frac{d\Gamma(B \to He^+e^-)}{dq^2}} \tag{1}$$

where H is either K or K^* meson and q^2 is the dilepton squared mass. These ratios are useful to test lepton flavor universality violation (LFUV) and within appropriately chosen ranges of q^2, these observables can be predicted very precisely in the SM; see [8,9] for details. The SM predictions are, respectively, $R(K) = 1.0004(8)$, and

$$R_{K^*} = \begin{cases} 0.920 \pm 0.007, & q^2 \in [0.045, 1.1] \text{ GeV}^2, \\ 0.996 \pm 0.002, & q^2 \in [1.1, 6] \text{ GeV}^2. \end{cases} \tag{2}$$

The LHCb collaboration has measured [10,11]

$$R_K = 0.846^{+0.060+0.016}_{-0.054-0.014}, \qquad q^2 \in [1.1, 6] \text{ GeV}^2, \tag{3}$$

and

$$R_{K^*} = \begin{cases} 0.660^{+0.110}_{-0.070} \pm 0.024, & q^2 \in [0.045, 1.1] \text{ GeV}^2, \\ 0.685^{+0.113}_{-0.069} \pm 0.047, & q^2 \in [1.1, 6] \text{ GeV}^2. \end{cases} \tag{4}$$

We will use the notation $R_{K^*}^{\text{Low}}$ and $R_{K^*}^{\text{Central}}$ from now on to represent R_{K^*} for values of q^2 in $[0.045, 1.1]$ GeV^2

*bhattacharyasrimoy@gmail.com
†iluvnpur@gmail.com
‡soumitra.nandi@iitg.ernet.in
§sunando.patra@gmail.com

and $[1.1, 6]$ GeV^2, respectively. Very recently Belle has measured the observables $R_{K^{(*)}}$ and the measured values are given by [12]

$$R_{K^*} = \begin{cases} 0.52^{+0.36}_{-0.26} \pm 0.05, & q^2 \in [0.045, 1.1] \text{ GeV}^2, \\ 0.96^{+0.45}_{-0.29} \pm 0.11, & q^2 \in [1.1, 6] \text{ GeV}^2. \end{cases} \quad (5)$$

and [13]

$$R_K = 0.98^{+0.27}_{-0.23} \pm 0.06, \qquad q^2 \in [1, 6] \text{ GeV}^2, \quad (6)$$

These new measurements have larger uncertainties compared to those from LHCb, but the results are consistent with each other. Belle has also measured separate ratios like $R^0_{K^*}$ and $R^+_{K^*}$, but the associated uncertainties are quite large at the moment. On the whole, the deviation between data and SM predictions stand at the level of 2.5 to 3σ. Future measurements of these ratios with enough statistical significance would have the potential to discover NP unambiguously.

The observed discrepancies can be explained in various NP models. Different types of new physics interactions (like vector, scalar etc.) with different Lorentz structures may contribute to these decays and explain the data. A lot of work has already been done and it is a difficult task to quote all of them. We are more interested, in the present work, in a model independent analysis. There are a few related analyses available in the literature, which mainly focus on considering one or two operators at a time [8,14–31].

In this article, we have done a model independent analysis of the NP affecting the $b \to s\ell^+\ell^-$ decay modes. The operator basis is exactly the same as that given in Ref. [23]. We have considered all possible combinations of these operators and categorized them as independent models (scenarios). There are several models capable of describing the observed data and one is thus confronted with the problem of model selection.

In short, the problem of model selection is as follows: any model, used to represent certain observation, will almost never be exact; chances are, that some information will be lost due to the choice of that particular model [32,33]. Choosing a simplistic model with too few parameters can involve making unrealistically simple assumptions and lead to high bias, poor prediction, and consequent missed opportunities for insight. While simplistic models are not flexible enough to describe the sample or the population well, a model with a larger number of parameters can fit the observed data very well. Does this make it a better model? With too many parameters, we face the possibility of just fitting the noise in the data and losing sight of the important trends. The most general problem in model selection is thus the optimization of the parameters required to explain certain observation [34,35]. The motivation: there must be a happy medium somewhere.

The most generally applicable, powerful, and reliable method for model comparison (also computationally expensive) is "cross-validation"[36], which, in addition to testing the predictive power of the model, minimizes the bias and variance together by minimizing the mean-squared-error (MSE). On the other hand, penalized-like-lihood information criteria, such as the Akaike information criterion (AIC) [37], and the Bayesian information criterion (BIC; more aptly named as Schwarz information criterion) are widely used for model selection. AIC estimates the relative amount of information lost by a given model: the less the information lost by a model, the higher the quality of that model. For a detailed discussion on various model selection procedures and their relative performance with respect to cross-validation, Ref. [38]. In our earlier publications [39,40], we have used these criteria in the context of NP model selections in $b \to c\tau\nu_\tau$ decays.

Very recently, in a Bayesian analysis of $b \to s\ell^+\ell^-$ decays [27], an information criterion has been used. They have shown the use of a criterion closely related to DIC (deviance information criterion; the definition by Gelman *et al.* [41]) and BPIC (Bayesian predictive information criterion [42]) for model selection, which is not only ideal for samples from a Markov Chain Monte Carlo but is also asymptotically equivalent [43] to natural model-robust version of AIC.

AIC is easy to calculate in a frequentist analysis, which is not the case for Bayesian analyses. The main difference between that analysis and ours is that they created the model hierarchy by defining $\Delta IC = IC_{SM} - IC_{NP}$. As a result, the quality of a model is determined with respect to the SM, whereas in our case the best model is picked up first and the hierarchy is defined with respect to that. Still, similar to our findings, they have found that the case ΔC_9 ($C^{NP}_{9,\mu}$ in their paper) provides the optimal outcome for $B \to K^*\ell\ell$ transitions if we consider only the "moments" data for the angular observables, in addition to the new LFUV data.

In the present analysis, we use both AIC and cross-validation to pin down the best possible model(s), and find out how one can use both procedures in tandem to glean the most out of the data at hand.

The article is organized as follows: in Sec. II, we discuss the present experimental and theoretical status of the observables used in this analysis. To further motivate a global analysis, we discuss the prospect of NP models in explaining present data in Section III. Section IV discusses the detailed methodology of the statistical analysis, as well as model selection. We present our results in Sec. V and in Sec. VI we summarize.

II. BACKDROP

A. Experimental

We list the experimental results used in this analysis and the corresponding references below:

(i) Binned data on the angular observables related to the $B^0 \to K^{*0}\mu^+\mu^-$ decays have been obtained from refs. [1] (LHCb) and [2] (ATLAS).[1]

(ii) Binned data for the differential branching fraction for $B^0 \to K^{*0}\mu^+\mu^-$ have been obtained from Ref. [45] and that for $B^+ \to K^{*+}\mu^+\mu^-$ from [46]. Both of these are LHCb references.

(iii) Binned data on the angular observables for $B^+ \to K^+\mu^+\mu^-$ (A_{FB} and F_H) have been taken from Ref. [47] (CMS).

(iv) Binned data on the differential branching fraction for $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0\mu^+\mu^-$ reported by LHCb have been taken from Ref. [46].

(v) Binned data on the angular observables for $B_s \to \phi\mu^+\mu^-$ (LHCb) have been taken from Ref. [48].

(vi) The lepton flavor universality violating (LFUV) observables R_{K^*}, both for the low and central bin, have been obtained from Ref. [10] (LHCb). We also include the recent measurements on these observables (for the same bins) from Belle [12]. The old R_K data from LHCb has been taken from Ref. [49]. The updated result on the same has also been included [11].

(vii) The experimental result for the branching ratio (BR) corresponding to $B_s \to \mu^+\mu^-$ has been taken from [50] which is the average of the measured values by CMS, ATLAS, and LHCb. The value is given by

$$\mathrm{Br}(B_s \to \mu\mu) = (3.1 \pm 0.6) \times 10^{-9}. \qquad (7)$$

The decay constant is taken from Refs. [51,52]

$$f_{B_s} = 0.2284 \pm 0.0037 \text{ GeV}. \qquad (8)$$

All numerical uncertainties quoted in this analysis, unless otherwise specified, denotes the 1-σ (68% c.l.) range.

A few words regarding the data on the angular observables due to LHCb taken from Ref. [1] is in order at this point. LHCb has provided the data corresponding to the angular observables in bins of q^2 ($q = p_{\mu^+} + p_{\mu^-}$, p_μ being the four-momentum of muon) by performing two separate analyses. The more commonly used dataset in the community is that due to the "method of moments." The angular observables in this case are determined by using a principal moment analysis of the angular distribution without carrying out any angular fit to the data [53,54]. These moments are continuous functions of q^2. The statistical uncertainties for these angular moments are estimated using a bootstrapping technique [55] and confidence intervals are defined such that they include the 16th–84th percentiles of the bootstrap distribution of the observables. The other

method termed the "maximum likelihood fit" involves an unbinned maximum likelihood fit to the invariant mass $m(\mu^+\mu^-(K^* \to)K^+\pi^-)$ and the three decay angles $\cos\theta_l$, $\cos\theta_K$ and ϕ in each q^2 bin, where:

(i) θ_l is the angle between the μ^+ (μ^-) and the direction opposite to that of the B^0 ($\bar{B}^0$) in the rest frame of the dimuon system,

(ii) θ_K is the angle between the direction of the K^+ (K^-) and the B^0 ($\bar{B}^0$) in the rest frame of the K^{*0} ($\bar{K}^{*0}$) where the K^{*0} meson is reconstructed through the decay $K^{*0} \to K^+\pi^-$, and

(iii) ϕ is the angle between the plane defined by the dimuon pair and the plane defined by the K and the π in the B^0 ($\bar{B}^0$) rest frame.

The bin sizes corresponding to the maximum likelihood analysis are larger than those for the method of moments. This is done since there is a dearth of statistics, and an increase in the bin-size renders the precision comparable with that for the method of moments. With increased number of events in the future, an unbinned likelihood analysis will become the norm, but at the present precision level, the moments data is at least equally dependable, if not more. To examine and point to any fundamental difference between these two datasets in presence of NP models, we use both these sets as separate cases in our analysis. To the best of our knowledge, this is the first global $b \to s\ell\ell$ analysis that takes both of these datasets into account.

Apart from classifying the data according to whether it corresponds to the "likelihood" or the "moment" method for the angular observables, we have also prepared separate datasets which we call:

(i) "Old" dataset, containing the (previous) estimates for the LFUV R_K and R_{K^*} ratios from Refs. [10,49] respectively, and

(ii) "New" dataset, where the old estimate for $R(K)$ [49] by LHCb is replaced by the new one [11], while both the previous [10] and the current [12] estimates for the R_{K^*} ratio have been included. We also include the most recent measurement for R_K due to Belle [13].

We should mention here that we have only taken the "low bins" ($q^2 \le 6$ GeV^2) corresponding to the experimental data referred to above. This is done so that we can avoid the region around the J/ψ resonance (the "broad charmonium" region) since a trustworthy theoretical estimate for this region is challenging. Hence, we do not include the R_{K^*} data from the low-recoil region provided by the recent Belle measurement from [12]. We take care of the systematic and statistical correlations separately in the data as and when they have been reported.

B. Theoretical

The effective Hamiltonian and the operator basis for exclusive $b \to s\mu^+\mu^-$ decays are taken from [22,23] and is written as:

---

[1]We refrain from using the very old (2012) CDF data available from the public note [44] on the angular observables in the $B \to K^{(*)}\mu^+\mu^-$ decays.

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}}(\lambda_t \mathcal{H}_{\text{eff}}^{(t)} + \lambda_u \mathcal{H}_{\text{eff}}^{(u)}) \qquad (9)$$

with the CKM combination $\lambda_i = V_{ib}V_{is}^*$ and

$$\mathcal{H}_{\text{eff}}^{(t)} = C_1 \mathcal{O}_1^c + C_2 \mathcal{O}_2^c + \sum_{i=3}^{6} C_i \mathcal{O}_i$$
$$+ \sum_{i=7,8,9,10,P,S} (C_i \mathcal{O}_i + C_i' \mathcal{O}_i'),$$

$$\mathcal{H}_{\text{eff}}^{(u)} = C_1(\mathcal{O}_1^c - \mathcal{O}_1^u) + C_2(\mathcal{O}_2^c - \mathcal{O}_2^u). \qquad (10)$$

We consider NP effects in the following operators:

$$\mathcal{O}_7 = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}, \quad \mathcal{O}_7' = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu},$$

$$\mathcal{O}_9 = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu\mu), \quad \mathcal{O}_9' = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\mu}\gamma^\mu\mu),$$

$$\mathcal{O}_{10} = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu\gamma_5\mu), \quad \mathcal{O}_{10}' = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\mu}\gamma^\mu\gamma_5\mu),$$

$$\mathcal{O}_S = \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{\mu}\mu), \quad \mathcal{O}_S' = \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{\mu}\mu),$$

$$\mathcal{O}_P = \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{\mu}\gamma_5\mu), \quad \mathcal{O}_P' = \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{\mu}\gamma_5\mu). \qquad (11)$$

The NP contributions to operators $\mathcal{O}_{9,10}$ is given by $\Delta C_{9,10}$. In these decays, when the final state contains a vector meson, one can construct various helicity amplitudes. These helicity amplitudes are used to form angular coefficients which are relevant in defining the *CP*-symmetric and asymmetric observables measured by the different experimental collaborations. The details about various transversity amplitudes and the respective angular coefficients can be obtained from [23]. The two major components that go into the formation of the helicity amplitudes are the Wilson coefficients (WC) of different operators and the form factors which are defined as the hadronic matrix elements of various operators. We follow Ref. [6] for the form factors in $B \rightarrow K^*$ and $B_s \rightarrow \phi$ decays.[2]

For the $B \rightarrow K$ sector we closely follow the methodology communicated in Ref. [57]. This includes expressing the differential decay distribution in terms of a polynomial in $\cos\theta$, where θ denotes the angle between the direction of motion of the parent B meson and the positively charged lepton in the dilepton center of mass frame. The coefficients of these terms can then be expressed as combinations of the corresponding WC and form factors. For the form

TABLE I. SM values and experimental world averages of inclusive and exclusive $b \rightarrow s\gamma$ observables used in our analysis.

Observable	SM prediction $\times10^5$	Measurement $\times10^5$
$\text{BR}(B \rightarrow X_s\gamma)_{E_\gamma > 1.6 \text{ GeV}}$	33.6 ± 2.6 [63]	32.7 ± 1.4 [64]
$\text{BR}(B^+ \rightarrow K^*\gamma)$	3.51 ± 0.78	4.21 ± 0.18 [65]
$\text{BR}(B^0 \rightarrow K^*\gamma)$	3.49 ± 0.78	4.33 ± 0.15 [65]
$\overline{\text{BR}}(B_s \rightarrow \phi\gamma)$	4.33 ± 0.77	3.5 ± 0.4 [66,67]

factors, we use the results from Ref. [22], where the authors perform a combined fit to the lattice computation in Ref. [58] as well as LCSR predictions at $q^2 = 0$ [59,60], using the parametrization and conventions of [58]. The method is described in details in the Appendix of Ref. [61].

We also take care of the correlations among these form factor elements as reported in these references, in order to propagate them to form the theoretical correlations and errors for the corresponding observables.

Since our aim is to perform a global model selection based on the plethora of available $b \rightarrow s\ell\ell$ data discussed in Sec. II A, there is a possibility that among the selected models the operator with C_7' as coefficient may appear as a plausible solution. Such an operator is also relevant for the radiative decays like inclusive and exclusive $b \rightarrow s\gamma$. For such scenarios, we have checked whether parameter spaces which are allowed by $b \rightarrow sll$ data are also allowed by the inclusive $B \rightarrow X_s\gamma$ measurement, alongwith the branching ratios for the three exclusive radiative modes $B^+ \rightarrow K^{*+}\gamma$, $B^0 \rightarrow K^{*0}\gamma$ and the time integrated $\overline{\text{BR}}(B_s \rightarrow \phi\gamma)$.[3] The definitions and formulas for these modes are taken from Ref. [62]. We provide the experimental values and the SM estimates used in our analysis in Table I. The corresponding theoretical (for $B \rightarrow X_s\gamma$) and experimental references are provided therein. Our SM values are consistent with the estimates of Ref. [62], within 1σ.

III. NP and current data on R_K and R_{K^*}

Before pursuing a detailed discussion on model selection, let us look for the NP effects in $b \rightarrow s\mu\mu$ decays, only in the light of recently updated measurements on R_K and R_{K^*} in this section, focusing on the measurements of R_{K^*} in the low q^2 bins. There is some discrepancy between this particular data and the corresponding predicted value in the SM. However, one needs to remember that the angular observables are not free from hadronic uncertainties. In this part of the study, we do not consider any of the angular observables, neither do we carry out any fit to data. We simply check the dependencies of $R_{K^*}^{\text{Low}}$, $R_{K^*}^{\text{Central}}$, and R_K on various WCs in one and two operator scenarios. We do

---

[2]Although the latest LCSR estimates for the $B \rightarrow K^*$ decays is calculated in Ref. [56], it does not include the corresponding $B_s \rightarrow \phi$ matrix elements. Hence we refrain from using these results.

[3]We refrain from using measurements for the *CP* asymmetries since our NP Wilson coefficients are taken to be real, thus excluding the possibility of *CP* violation in NP.
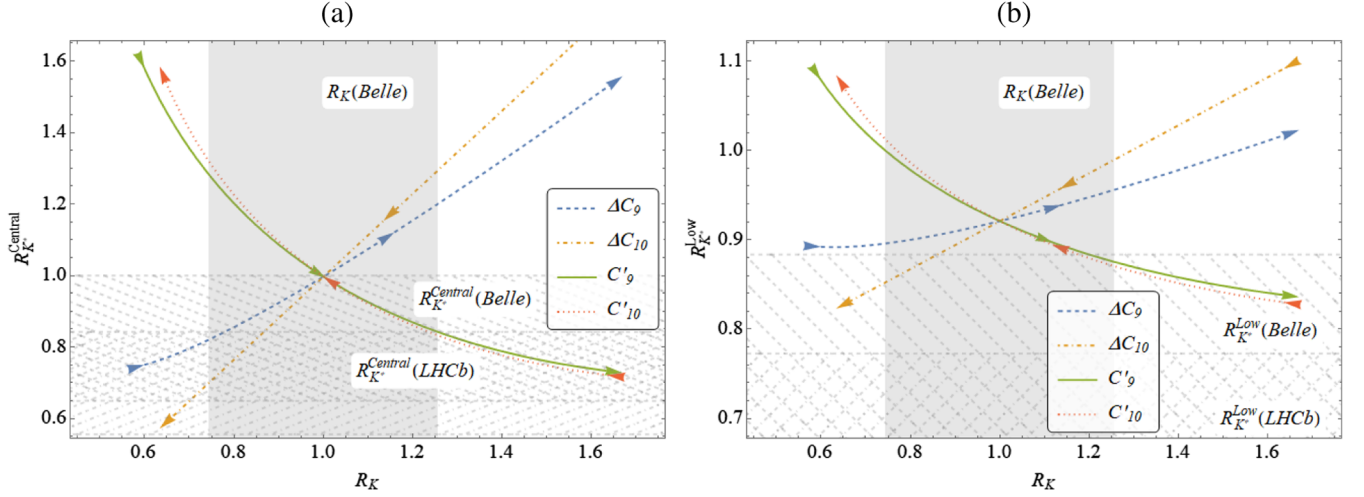
FIG. 1.   Correlations between R_K and R_{K^*} for different single-operator NP scenarios. The arrows indicate the increasing values of the WCs from -2 to $+2$. All the experimental data are considered within their 1σ ranges.

not include C_S, C_P, C'_S, and C'_P, since the corresponding operators by themselves, or combinations including such operators, are unable to explain the observed data in $R_{K^*}^{\text{Low}}$. These WCs also suffer from tight constraints due to $B_s \to \mu\mu$ decays [68]. Also, the new electromagnetic dipole operator $\mathcal{O}'_7$ alone would not be able to explain the observed data on $R_K^{(*)}$ and the branching fractions in the above mentioned radiative decays simultaneously. Hence, we have not considered the effects of this operator in this part of the analysis.

The results of our analysis are presented in Figs. 1 and 2. In single operator scenarios, the correlations between all the above-mentioned observables are shown in Figs. 1(a) and 1(b). It would be difficult to explain the observed data for $R_{K^*}^{\text{Low}}$ within their 1σ ranges will be difficult in the single-operator scenarios. Although the allowed region is tightly constrained, $\mathcal{O}_{10}$ (with WC ΔC_{10}) is the only operator that can simultaneously explain all the data on $R_{K^{(*)}}$ except $R_{K^*}^{\text{Low}}$ from LHCb. The required value of ΔC_{10} lies in between 0.5 and 1.5, which is consistent with the measured value of $\text{Br}(B_s \to \mu\mu)$ within its 2-σ range for detail see Fig. 2(k). However, there are several candidates in the two operator scenarios that could explain all the data simultaneously. Among various possible combinations, the highly probable scenarios are the operators with the WCs $[\Delta C_9, C'_{10}]$, $[\Delta C_{10}, C'_9]$, $[\Delta C_{10}, C'_{10}]$, and $[\Delta C_{10}, \Delta C_9]$. The other possible scenario $[\Delta C_9, C'_9]$ is less favored but allowed by the data. Also, the allowed values of ΔC_{10} and/or C'_{10} can explain $\text{Br}(B_s \to \mu\mu)$ within its 1-σ range, wherever applicable; see Fig. 2(k) for details. In Fig. 2(k), we have shown the variations of $\text{Br}(B_s \to \mu\mu)$ with respect to the parameters $[\Delta C_{10}, C'_{10}]$. However, there are scenarios where only ΔC_{10} or C'_{10} appears. In such cases, depending on the scenario, one needs to look at the plot with either $\Delta C_{10} = 0$ or $C'_{10} = 0$. To conclude this section, we would

like to mention that the three or more operator scenarios could also be relevant to explain the present data on $R_{K^{(*)}}$ simultaneously. The take home message is that simultaneous contributions from various operators are required for a simultaneous explanation of the $R_{K^{(*)}}$ data alone. The results of this section will be useful for a better understanding of the results in the following section.

IV. METHODOLOGY

A. Parameter estimation

The methodology adopted in this paper for the model selection is as follows:

(1) *Define models*: Considering the NP Wilson coefficients real, we take all possible combinations (511 in total) of the coefficients forming a predefined global set of different scenarios. Each scenario with a specific combination of coefficients thus constitutes a potential model to explain the experimental results.

(2) *Numerical optimization*: Next, for each such model k, as mentioned above, we perform a frequentist statistical analysis optimizing a χ^2 statistic which is a function of the Wilson coefficients. Whenever applicable, statistical (systematic) covariance matrices $V^{\text{stat(syst)}}$, are constructed by taking separate correlations. Theoretical uncertainties are propagated separately and are introduced in the χ^2 in terms of a "theoretical" covariance matrix V^{th}. The effect of the interplay of the SM uncertainties and the NP parameters come in the fit at a higher order and are neglected without any loss of generality. Following Sec. II A, we perform 4 types of fit for each model:
(a) *New* data with *Likelihood* data for angular observables, a total of 214 observables.
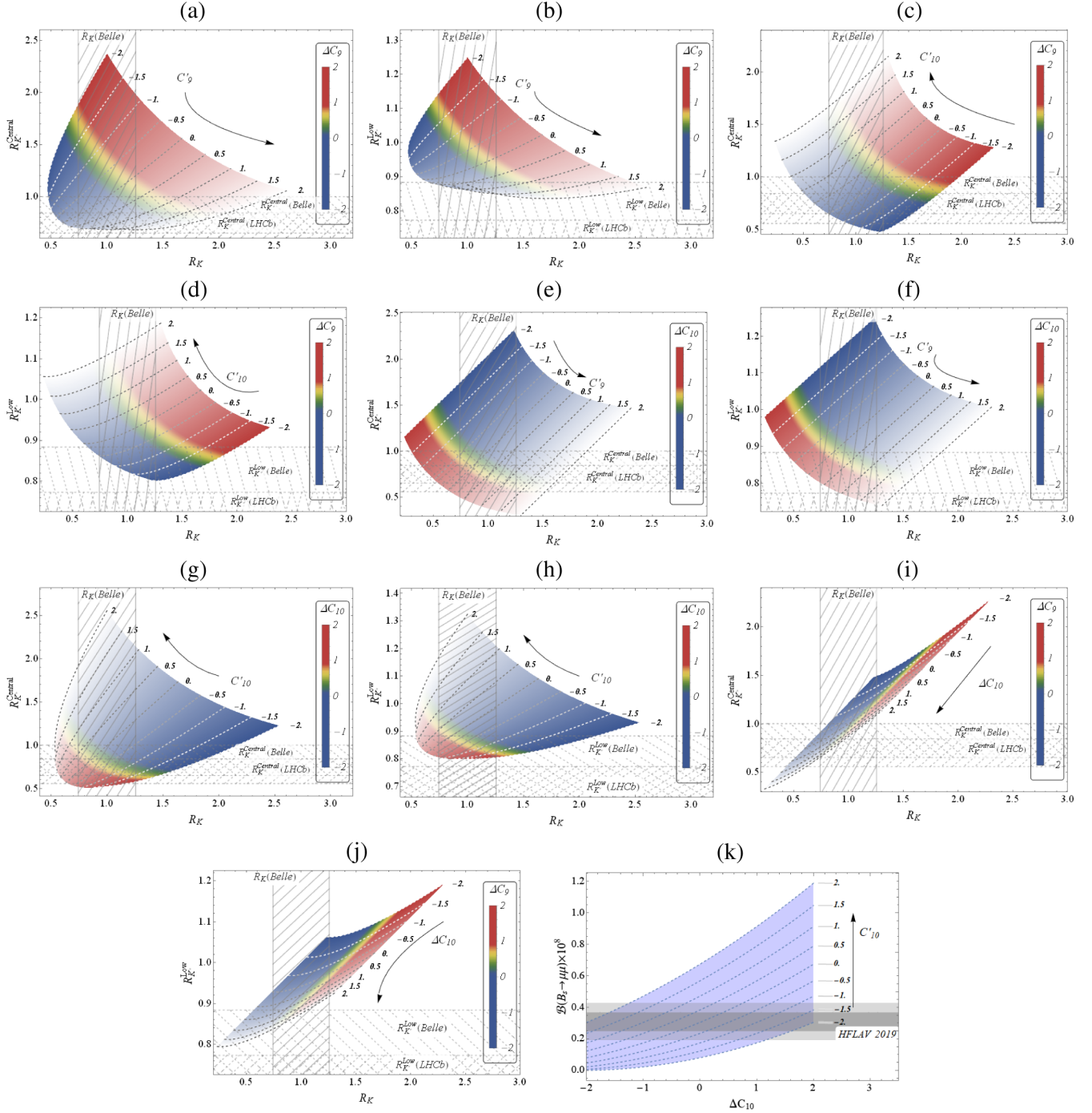
FIG. 2. Figure 2(a) to 2(j) shows the correlations between R_K and R_{K^*} in different NP scenarios. The constraints on ΔC_{10} and C'_{10} from the measured value of $\mathrm{Br}(B_s \to \mu\mu)$ can be inferred from Fig. 2(k).

(b) *New* data with *Moments* data for angular observables, a total of 258 observables.

(c) *Old* data with *Likelihood* data for angular observables, a total of 211 observables.

(d) *Old* data with *Moments* data for angular observables, a total of 255 observables.

All fits are done in batch using *Mathematica©* in the form of a package [69]. The chosen optimization method is

differential evolution, a stochastic parallel direct search evolution strategy [70].[4]

(3) *Postprocess*: In the postprocess for each fit, we obtain fit-quality using *p*-value and find outliers by

---

[4]Capable of handling nondifferentiable, nonlinear, and multimodal objective functions. Considerably faster than generic genetic algorithms and extremely able to find the global minimum.

constructing a pull (related to studentized residuals; for our purpose, the difference between the fitted and experimental results, normalized by the uncertainty of the data, including theory uncertainties [71,72]) for each data-point. We also check the normality of the pull-distribution (i.e., consistency with a Gaussian of $\mu = 0$ and $\sigma = 1$) to ensure the applicability of the χ^2 as the fit-statistic. We use the "Cramér-von Mises" criterion [73] for the normality check. Scenarios not satisfying the normality criterion are dropped from the analysis.

(4) *Parameter-space*: Parameter uncertainties are obtained both from the Fisher matrix[5] and the profile-likelihood curve[6]

With the remaining scenarios, we perform a model-selection procedure for each dataset. In the following subsection, we elaborate the methods used to do the multimodel selection procedure.

B. Model selection

Following the "concept of parsimony" [76], we need to optimize the dimension (measure of the degree of structure) of the model explaining our data. All model selection methods, to some extent, depend on the principle of parsimony [77]. In statistical terms, this is expressed as a bias versus variance trade-off. In general, bias decreases and variance increases as the model-dimension increases.

1. Cross-validation

As we have mentioned in the Introduction, cross-validation is the most generally applicable, powerful, reliable, and computationally expensive method for model comparison. The most straightforward and the most expensive flavor of cross-validation is "leave-one-out cross-validation" (LOOCV). In LOOCV, one of the data points is left out and the rest of the sample ("training set") is optimized for a particular model. Then that result is used to find the predicted residual for the left out data point. This process is repeated for all data points and a mean-squared-error (MSE) is obtained using all those residuals. This process is repeated for all models. The models with the least MSE are the best ones.

2. Criteria from information theory

In addition to the extreme computational cost demanded by cross-validation methods, especially LOOCV, its

applicability is questionable to very small sample sizes [78,79]. Due to this reason, in our earlier works [39,40], we have shown the importance and use of the information-theoretic criterion AIC [37] and its second order variant AIC_c [80]. It has been shown that minimizing AIC is asymptotically equivalent to cross-validation [81]. For a detailed discussion on AIC_c, we point the reader to those papers and references therein. Here, let us reiterate the main important aspects of AIC_c in with respect to model selection in the present work:

(a) AIC_c: If the full reality or truth is noted as f and an approximating model in terms of probability distribution is g, then we can define a model selection criterion in terms of the $\chi^2_{\min}$ (equivalent to the maximum point of the empirical log-likelihood function) in the parameter space:

$$AIC_c = \chi^2_{\min} + 2K + \frac{2K(K+1)}{n-K-1} \qquad (12)$$

where n is the number of data points and K is the number of estimable parameters.[7]

(b) $w_i^{\Delta AIC_c}$: The model which is the closest to the unknown reality generating the data should have the smallest value of AIC_c among the considered models. Simple differences of them ($\Delta_i^{AIC} = AIC_c^i - AIC_c^{\min}$) estimate the relative expected information loss between f and g_i allowing comparison and ranking of candidate models in increasing order of Δ_i^{AIC}. Generally, the level of empirical support in favor of g_i is considered substantial when ΔAIC_c is between 0 and 2 ($\Delta AIC_c \leq 4$ is considered to be a conservative and loose bound). We can also quantify the weight of evidence in favor of model i by defining a set of positive "Akaike weights":

$$w_i^{\Delta AIC_c} = \frac{e^{(-\Delta_i^{AIC}/2)}}{\sum_{r=1}^{R} e^{(-\Delta_r^{AIC}/2)}}, \qquad (13)$$

adding up to 1 [34]. As these depend on the entire set, adding or dropping a model during an analysis requires recomputation for all models in the new set.

In the present analysis, we have a unique opportunity to not only test the relative capability of MSE from cross-validation and $w_i^{\Delta AIC_c}$, but also the validity of the empirical rule of selecting models with $\Delta AIC_c \leq 2$. To that end, we first select a large number of competing models by

---

[5]In case of approximately Gaussian parameter-profile likelihoods, it is possible to obtain the HESSE errors [74], which are, obviously, symmetric.

[6]Range of the 1σ confidence level (CL) of the profile likelihoods of the said parameter. One and two dimensional profile likelihoods in this analysis will be depicted as 1-CL plots, closely following the PROB method followed in Ref. [75]

[7]A more preferable way of calculating "number of estimable parameters" is to calculate the p-value of the fit from toy Monte-Carlo (MC) method. Under the assumption that the fit-statistic follows a χ^2 distribution, this can give us the number of degrees of freedom, and thus the number of estimable parameters. Still, as we need the differences between the AIC_c values instead of the absolute ones, the naive way of parameter counting works fine.
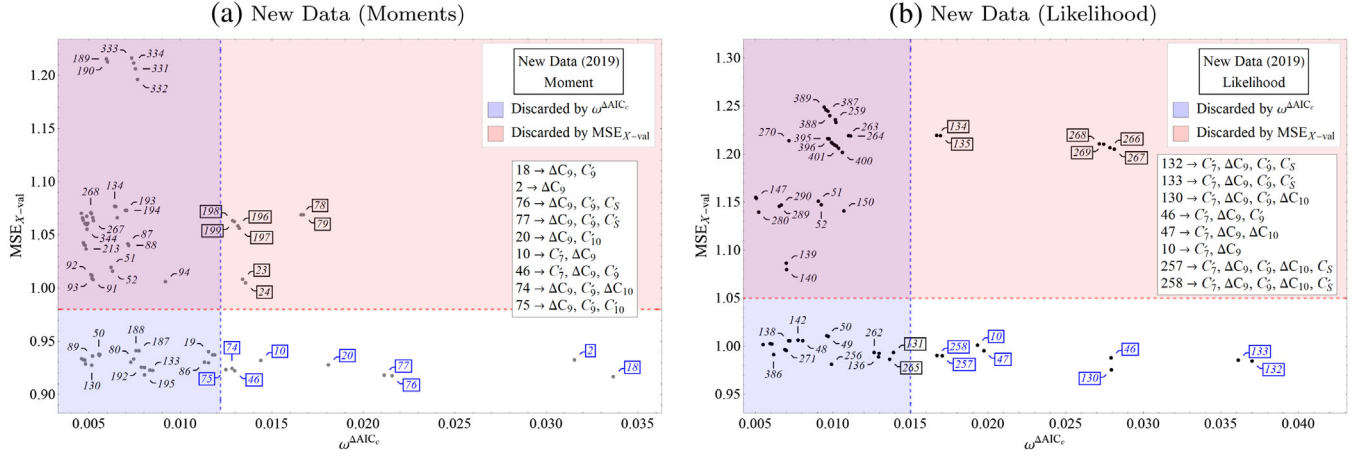
FIG. 3. For the fit with "New Data," indices of competing scenarios with $\Delta AIC_c \leq 4$ in the $MSE_{X\text{-val}}$ vs $w_i^{\Delta AIC_c}$ plane. We break the plane in four regions, with the one in right-bottom being the best one. Models in this region are chosen as the best ones from these two criteria and the labels are colored blue. Wilson coefficients contained in these models are shown in-box. For comparison, indices for models picked up by the criterion $\Delta AIC_c \leq 2$ are framed. For details, check Sec. V A.

using the conservative limit of $\Delta AIC_c \leq 4$, and then distribute them in the plane of $MSE_{X\text{-val}}$ vs $w^{\Delta AIC_c}$ and check how they are clustered. Models with a low value of $MSE_{X\text{-val}}$ and a high value of $w^{\Delta AIC_c}$ are the undoubtedly the best ones to explain the data.

V. RESULTS

A. Model selection

As explained in Sec. IV, we perform the fit for four different sets. After applying the normality check (as explained in Sec. IV A) on all the 511 models (thus ensuring only valid fits remain in our dataset), we pick out the large set of models with $\Delta AIC_c \leq 4$. The list of models, thus selected, are shown in Figs. 3 and 4, which are based on the analysis of all the available data sets given in Sec. II A. Each point in these figures represents a "selected" model (i.e., a model for which $\Delta AIC_c \leq 4$). Among these,

the indices (labels for points) for models selected by $\Delta AIC_c \leq 2$ are framed. As is evident from the figure (and explained earlier in Sec. IV B 1), the lower the $MSE_{X\text{-val}}$, the better the model. One can clearly see three separate clusters depending only on $MSE_{X\text{-val}}$, and we can safely label the lowest one as the cluster of the best models (from $MSE_{X\text{-val}}$) and discard the rest. Similarly, there are three clusters in the $w^{\Delta AIC_c}$ direction as well, where the cluster with the largest $w^{\Delta AIC_c}$ value contains only one model. All models with $\Delta AIC_c \leq 2$ lie in the two rightmost clusters. Following Sec. IV B 2, we know that a model with a larger value of $w^{\Delta AIC_c}$ is more probable to explain the data. So, we put a commensurate bound on $w^{\Delta AIC_c}$. We note that out of the various possible combinations, only a few are "selected" by the combined criteria of $MSE_{X\text{-val}}$ and AIC_c.

Figures 3(a) and 3(b) respectively compare the selected models in the analyses of moments and likelihood data on angular observables along with the new data on $R_{K^{(*)}}$.
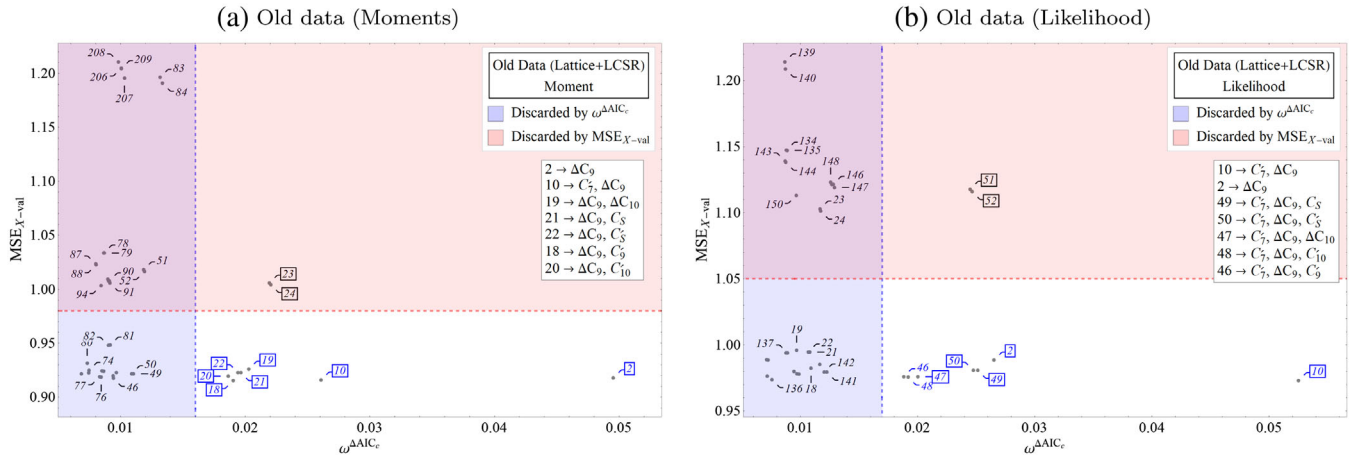

FIG. 4. Same as Fig. 4, but for the fit with "Old Data."

TABLE II. Fit-qualities, model selection criteria, parameter estimates and effects on radiative decays for the "best" selected models with the "new" dataset, with the "moments" estimate of the angular observables. Selected models are obtained from Fig. 3a. Last four columns showcase the deviations (in units of σ) between the experimental value of the radiative decays and the corresponding value obtained with the fit results.

Model index	χ^2_{Min}/DOF	p-val (%)	$\omega^{\Delta \text{AIC}_c}$ (%)	$\text{MSE}_{X\text{-val}}$	Parameter values	Deviation in σ			
						$B \rightarrow X_s\gamma$	$B^+ \rightarrow K^*\gamma$	$\Delta B^0 \rightarrow K^*\gamma$	$\Delta B_s \rightarrow \phi\gamma$
18	250.28/256	58.9	3.4	0.917	$\Delta C_9 \rightarrow -1.13 \pm 0.13$ $C_9' \rightarrow 0.25 \pm 0.17$	–	–	–	–
2	252.44/257	56.9	3.2	0.933	$\Delta C_9 \rightarrow -1.12 \pm 0.13$	–	–	–	–
76	249.12/255	59.2	2.2	0.918	$\Delta C_9 \rightarrow -1.18 \pm 0.14$ $C_9' \rightarrow 0.34 \pm 0.19$ $C_S \rightarrow -0.035 \pm 0.016$	–	–	–	–
77	249.16/255	59.1	2.1	0.918	$\Delta C_9 \rightarrow -1.18 \pm 0.14$ $C_9' \rightarrow 0.34 \pm 0.19$ $C_S' \rightarrow 0.035 \pm 0.016$	–	–	–	–
20	251.52/256	56.7	1.8	0.928	$\Delta C_9 \rightarrow -1.15 \pm 0.14$ $C_{10}' \rightarrow -0.1 \pm 0.104$	–	–	–	–
10	251.97/256	55.9	1.4	0.932	$C_7' \rightarrow 0.01 \pm 0.015\Delta$ $C_9 \rightarrow -1.15 \pm 0.14$	0.31	−0.87	−1.06	1.22
46	250.14/255	57.4	1.3	0.922	$C_7' \rightarrow 0.0058 \pm 0.0155$ $\Delta C_9 \rightarrow -1.15 \pm 0.14$ $C_9' \rightarrow 0.24 \pm 0.18$	0.3	−0.87	−1.06	1.22
74	250.16/255	57.4	1.3	0.925	$\Delta C_9 \rightarrow -1.16 \pm 0.15$ $C_9' \rightarrow 0.26 \pm 0.17$ $\Delta C_{10} \rightarrow -0.041 \pm 0.118$	–	–	–	–
75	250.21/255	57.3	1.2	0.923	$\Delta C_9 \rightarrow -1.12 \pm 0.14$ $C_9' \rightarrow 0.3 \pm 0.26$ $C_{10}' \rightarrow 0.04 \pm 0.157$	–	–	–	–

Similar comparisons are done with the old data on $R_{K^{(*)}}$ in Figs. 4(a) and 4(b), respectively. We note that apart from the two-operator scenario $[\mathcal{O}_7, \mathcal{O}_9]$, the likelihood and moments data pick up very different combinations of operators after model selection. The likelihood data prefer scenarios with more operators than those that are required for moments data. This could be since most of the observables which are determined from an unbinned maximum likelihood fit are consistent with their respective SM predictions. A large value of a single WC may lead to a discrepancy between the measured values and the corresponding SM predictions. It is thus preferable to have simultaneous contributions from different operators. The same observation can be made from the comparison of the quality of fits with moments and likelihood data which are shown in Tables II and III, respectively. The values of the reduced χ^2 and corresponding p-values, indicating the quality of the fit and the relative quality of the various models (scenarios) for a given set of data are provided in each corresponding table. We note that in general, the quality of the fits is better in the analysis with moments data

compared to that with likelihood data. It can be seen that the ranking of the models depends not only on the quality of fit but also on the penalty function defined in Eq. (12). This is one of the major advantage of AICc over data-fitting. Results of a similar analysis after dropping $R_{K^{(*)}}$ are presented in Fig. 9 in Appendix B. We note that after dropping the LFUV observables from the fits for the respective cases, the multioperator scenarios cannot survive the competition from those scenarios with relatively less number of operators.

As we know, the data from Belle has significant errors compared to those given by LHCb. To check whether this data has any influence over the selected scenarios, we have performed an analysis where we drop the $R_{K^{(*)}}$ data from Belle only. We note that the selected scenarios are the same as that given in Fig. 3. Also, the quality of fits or the respective p-values do not change. The reason is straightforward: the Belle data being consistent with LHCb, and the extracted $R(K)$ and $R(K^*)$ in the selected models in Fig. 3 are consistent with both LHCb and Belle. Hence, the conclusions remain unchanged even after dropping Belle data.

TABLE III.   Same as the Table II, but with the "likelihood" estimate of the angular observables. Selected models are obtained from Fig. 3(b).

Model index	χ^2_{Min}/DOF	p-val (%)	$\omega^{\Delta AIC_c}$ (%)	$MSE_{X\text{-val}}$	Parameter values	Deviation in σ			
						$B \to X_s\gamma$	$B^+ \to K^*\gamma$	$\Delta B^0 \to K^*\gamma$	$\Delta B_s \to \phi\gamma$
132	217.02/210	35.5	3.7	0.985	$C'_7 \to 0.04 \pm 0.015$ $\Delta C_9 \to -1.39 \pm 0.13$ $C'_9 \to 0.45 \pm 0.2$ $C_S \to -0.042 \pm 0.013$	0.44	−0.82	−1.02	1.27
133	217.07/210	35.4	3.6	0.986	$C'_7 \to 0.04 \pm 0.015$ $\Delta C_9 \to -1.39 \pm 0.13$ $C'_9 \to 0.45 \pm 0.2$ $C'_S \to 0.042 \pm 0.013$	0.44	−0.82	−1.02	1.27
130	217.58/210	34.5	2.8	0.976	$C'_7 \to 0.044 \pm 0.015$ $\Delta C_9 \to -1.42 \pm 0.14$ $C'_9 \to 0.32 \pm 0.19$ $\Delta C_{10} \to -0.16 \pm 0.11$	0.47	−0.81	−1.01	1.28
46	219.66/211	32.7	2.8	0.988	$C'_7 \to 0.04 \pm 0.015$ $\Delta C_9 \to -1.34 \pm 0.13$ $C'_9 \to 0.33 \pm 0.2$	0.44	−0.82	−1.02	1.27
47	220.36/211	31.5	2.	0.995	$C'_7 \to 0.048 \pm 0.015$ $\Delta C_9 \to -1.43 \pm 0.15$ $\Delta C_{10} \to -0.16 \pm 0.11$	0.5	−0.8	−1	1.29
10	222.46/212	29.7	1.9	1.001	$C'_7 \to 0.043 \pm 0.015$ $\Delta C_9 \to -1.33 \pm 0.13$	0.46	−0.81	−1.01	1.28
257	216.47/209	34.7	1.7	0.99	$C'_7 \to 0.042 \pm 0.015$ $\Delta C_9 \to -1.42 \pm 0.14$ $C'_9 \to 0.41 \pm 0.21$ $\Delta C_{10} \to -0.091 \pm 0.123$ $C_S \to -0.036 \pm 0.017$	0.46	−0.82	−1.01	1.28
258	216.51/209	34.6	1.7	0.99	$C'_7 \to 0.042 \pm 0.015$ $\Delta C_9 \to -1.42 \pm 0.14$ $C'_9 \to 0.41 \pm 0.21$ $\Delta C_{10} \to -0.092 \pm 0.123$ $C'_S \to 0.036 \pm 0.017$	0.46	−0.82	−1.01	1.28
131	218.98/210	32.1	1.4	0.993	$C'_7 \to 0.04 \pm 0.015$ $\Delta C_9 \to -1.32 \pm 0.13$ $C'_9 \to 0.47 \pm 0.25$ $C'_{10} \to 0.11 \pm 0.14$	0.44	−0.82	−1.02	1.27
265	216.92/209	33.9	1.4	0.986	$C'_7 \to 0.04 \pm 0.015$ $\Delta C_9 \to -1.39 \pm 0.13$ $C'_9 \to 0.44 \pm 0.2$ $C_S \to -0.24 \pm 0.6$ $C'_S \to -0.19 \pm 0.6$	0.44	−0.82	−1.02	1.27

There are a few other relevant observations here as well. For example, both in the case of the old and new datasets, the majority of the selected scenarios are with two, three, or four operators and all of them contain the operator corresponding to the WC ΔC_9. Also, the only scenario with a single operator selected by the new dataset (in particular the moments dataset) has the WC ΔC_9 (model 2). Model 2 is clearly the better option for all the datasets. However, for the new moments dataset, scenarios with two operators corresponding to the WCs $[\Delta C_9, C'_9]$ (model 18), $[\Delta C_9, C'_7]$ (model 10), and $[\Delta C_9, C'_{10}]$ (model 20) are other possible choices. An important feature of the data is that the

likelihood and moments datasets pick completely different types of models. For likelihood, an explanation of the observed data with a single operator is less plausible.

We note that some models have low MSE scores but are rejected by AICc. Although leave-one-out-cross-validation (LOO-CV) is asymptotically equivalent to AIC, there are differences between the two. Theoretical considerations aside, AIC is just likelihood penalized by the degrees of freedom. Evidently, AIC accounts for uncertainty in the data (-2Log(L)) and assumes that more parameters lead to a higher risk of overfitting (2k). Cross-validation (CV) just looks at the test set performance of the model, with no further assumptions. There is no explicit measure of model complexity, unlike AIC. Clearly, AIC penalizes model complexity more than CV. The accepted practice in the literature is that if one cares mostly about making predictions and assumes that the test set(s) to be reasonably

similar to the validation sets, one should go for CV (only with a large number of data) [38].

Following the reasoning in the paragraph above, it seems clear *a priori* that models with low MSE but not selected by AICc may have more parameters. In our analysis, the models are organized in such a way that more complex models with a larger number of parameters have a higher index. A quick examination of the plots in Figs. 3 and 4 validates this assumption, as the left sides of the plots are littered with models of higher indices.

On another note, as the title suggests, this study clearly finds that under the simultaneous application of both AIC and CV, the models cluster in such a way that lets us carry out a tighter selection of models than would have been possible by the use of any one of these criteria. So the fact that some of the selected models by CV are further discarded by AIC due to relative complexity is actually quite a nontrivial finding.
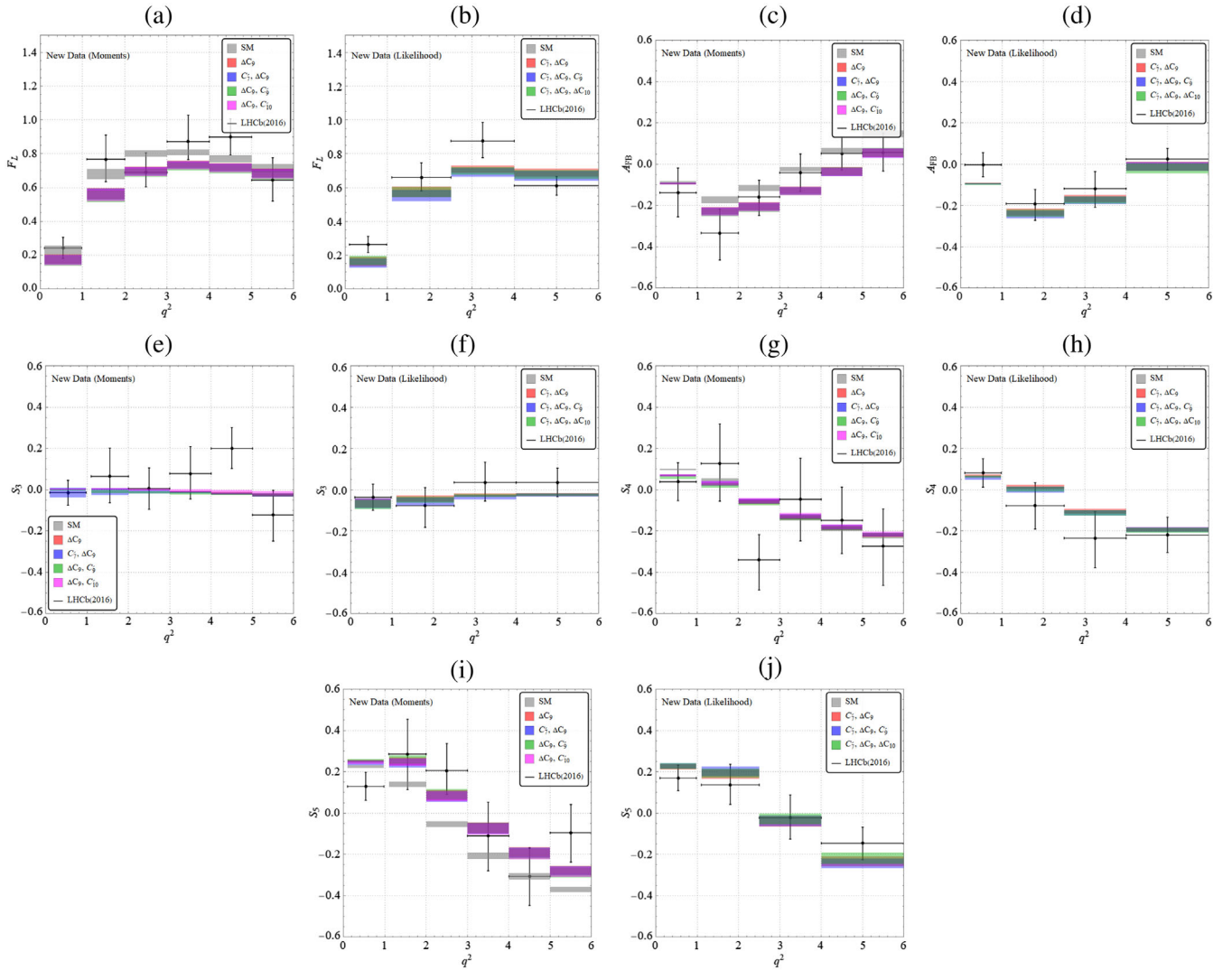

FIG. 5. Comparison of the *CP*-averaged angular observables obtained in experiment, SM and from our fit results considering all the available inputs.

The best fit values of the new WCs with the corresponding errors for the selected models in Fig. 3 are given in Tables II and III. We note that for all the fit results, the allowed values of ΔC_9 are negative and greater than one. Other WCs appearing along with ΔC_9 as probable solutions have values $\ll 1$ or -1. The selected models with C_7' as one of the WCs will impact radiative decays. Hence, along with the allowed values for the new WCs, we have shown that in all the relevant models the branching fractions for $B \to K^* \gamma$, $B \to \phi \gamma$ and $B \to X_s \gamma$ are consistent with their respective measurements within 1σ confidence interval in the relevant tables. The fitted values of the selected WCs remain almost unchanged in the analysis obtained after dropping $R_{K^{(*)}}$. The allowed parameter spaces of the respective WC's are similar in the analyses with old dataset which are given in the Appendix in Tables IV and V. However, note that the fit qualities are relatively better in the analysis with old data which is an indicator of relatively poor NP-description of the new LFUV observables along with the angular observables.

We compare our fit results for various angular observables with their respective measured values and the SM predictions in Figs. 5 and 6 for a set of selected models. We see that the data in a few bins are inconsistent with their corresponding SM predictions, particularly in the data set for the moment analysis. For the likelihood analysis, most of the measured values in different bins are consistent with their SM predictions, albeit with exceptions. Assuming the observed discrepancies are due to NP, most of them can be

resolved by our selected models. For the likelihood data, our analysis shows that the angular observables obtained from the selected models are fully consistent with their SM predictions as well as measured values. On the contrary, the fitted angular observables in the selected models in the moment-data analysis are shifted from their respective SM predictions in a few bins. Moreover, in some of those, the fit results shift from their measured values as well. One can also notice the correlations among various angular observables in the presence of different new operators from these figures. For example, A_{FB} and F_L are positively correlated, F_L and S_5 are negatively correlated, etc. For some of the bins of F_L, the values predicted by the selected models are shifted from respective measured values. Similar plots obtained in our analysis after dropping $R_{K^{(*)}}$ is provided in Fig. 10 in the Appendix.

In Fig. 7, the predicted values of different observables and their respective correlations are shown for a few of the models (only one or two operator scenarios) selected in our analysis. We note that $R_{K^*}^{\text{Central}}$ and R_K are correlated differently in different two operator scenarios. However, all of them satisfy the current experimental bounds on these observables. Interestingly, the single operator scenario with ΔC_9 and the two operator scenario with $[\Delta C_9, C_7']$ are unable to satisfy the current experimental bounds on $R_{K^*}^{\text{Low}}$. In these scenarios, satisfying the experimental bounds on R_K, it is hard to get a value of $R_{K^*}^{\text{Low}}$ below 0.85. However, the scenarios with $[\Delta C_9, C_9']$ and $[\Delta C_9, C_{10}']$ as WCs can explain the observed data on R_K and R_{K^*}. From the
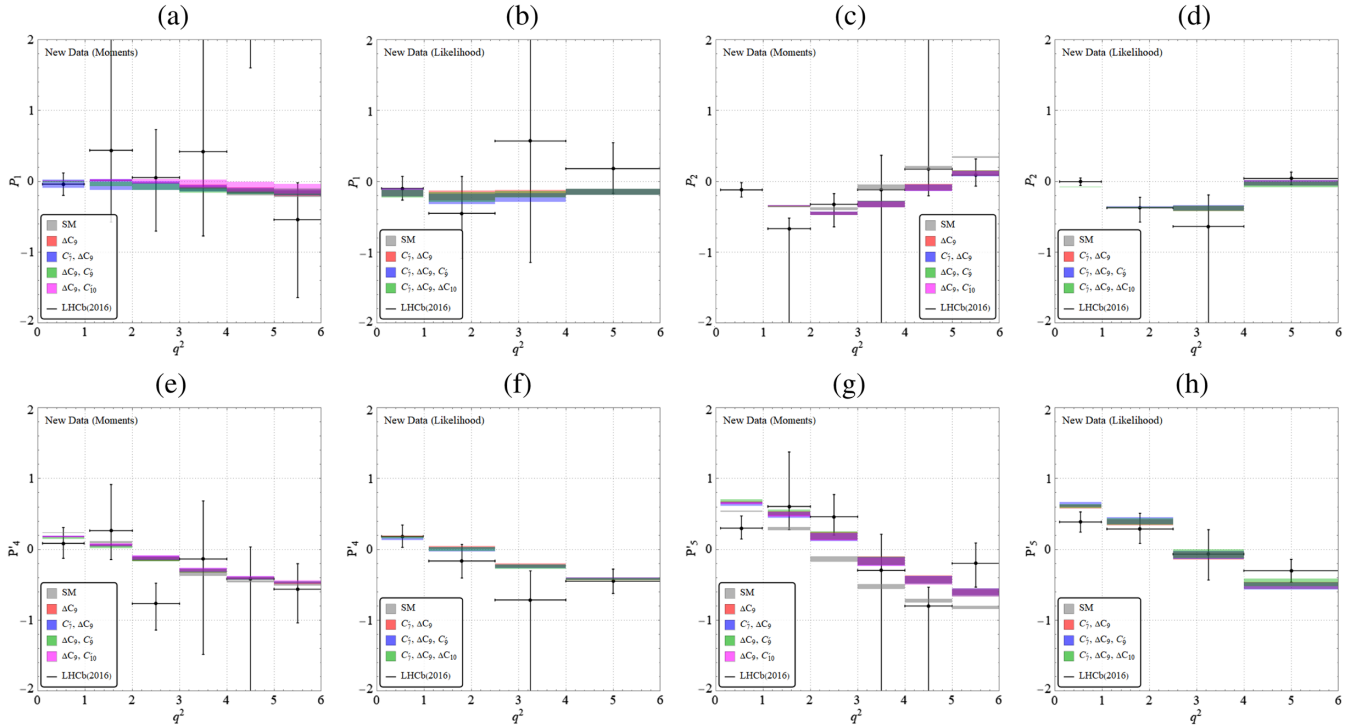

FIG. 6. Comparison of the optimized angular observables obtained in experiment, SM and from our fit results considering all the avilable inputs.
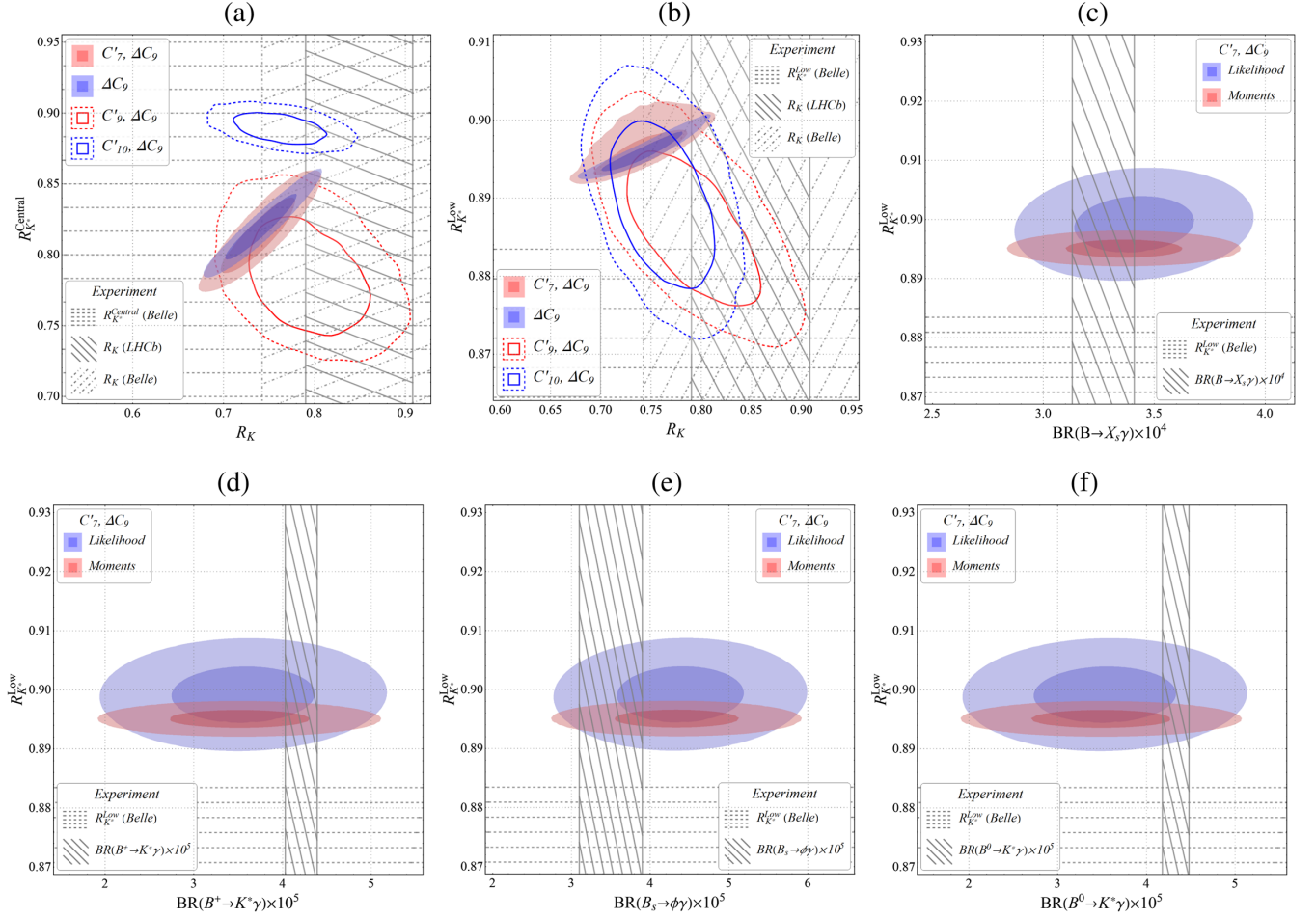
FIG. 7. Predicted values, and the correlation of different observables in a few selected scenarios.

respective correlations, one can also see that $R_{K^*}^{\text{Low}}$ and R_K are negatively correlated and values of $R_{K^*}^{\text{Low}}$ lower than 0.88 prefers $R_K > 0.8$. In Figs. 7(c), 7(d), 7(e), and 7(f), we provide the predicted values of the branching fractions of different radiative decays and their correlations with $R_{K^*}^{\text{Low}}$ in scenarios corresponding to the WCs $[\Delta C_9, C_7']$. We note that there are no noticeable correlations between these branching fractions and $R_{K^*}^{\text{Low}}$, or for that matter with R_K and $R_{K^*}^{\text{Central}}$.

The q^2 distributions and the zero crossing of the angular observables A_{FB}, S_4 and S_5 corresponding to the SM and the selected models are shown in Fig. 8. We have noted discrepancies between the q^2 distributions for the SM and the selected models in the cases of A_{FB} and S_5 while $S_4(q^2)$ in the selected models are fully consistent with the SM. The q^2 distributions of these angular observables in different selected models overlap with each other. Hence it is hard to discriminate models from these distributions. In the Appendix in Fig. 12, we compare the values of q^2 at the zero crossing (q_0^2) between SM, our selected models and the measured values for the above mentioned observables.

The uncertainties of the observables, in terms of the parameters, are obtained by two different techniques for Figs. 7 and 8. For Fig. 7, which contains spaces for any 2 observables, we take, from our fit-results, all the information (best-fit values, uncertainties, and correlations) of only the NP parameters occurring in those observables, create a multivariate distribution out of those, and sample a large number of points ($\sim$5000) from that distribution using Monte-Carlo. For each of those points, we get sets of values of the observables, which in turn lets us draw the $1\,(39.35\%)$ and $2\,(63.21\%)\sigma$ contours (for 2-observable plots) from these datasets.

For Fig. 8, we are dealing with only one observable at a time and that too, for a specific value of q^2. Taking the 1σ (68%) confidence levels for the marginal likelihoods around the central (for SM/nuisance parameters) or best-fit values (for NP) of the parameters as uncertainties and the corresponding correlations between them, we propagate the uncertainties to get the central values and uncertainties of the observables at some q^2. Doing this for many points over the allowed q^2 range and interpolating between them gives us the plots of q^2 distribution.
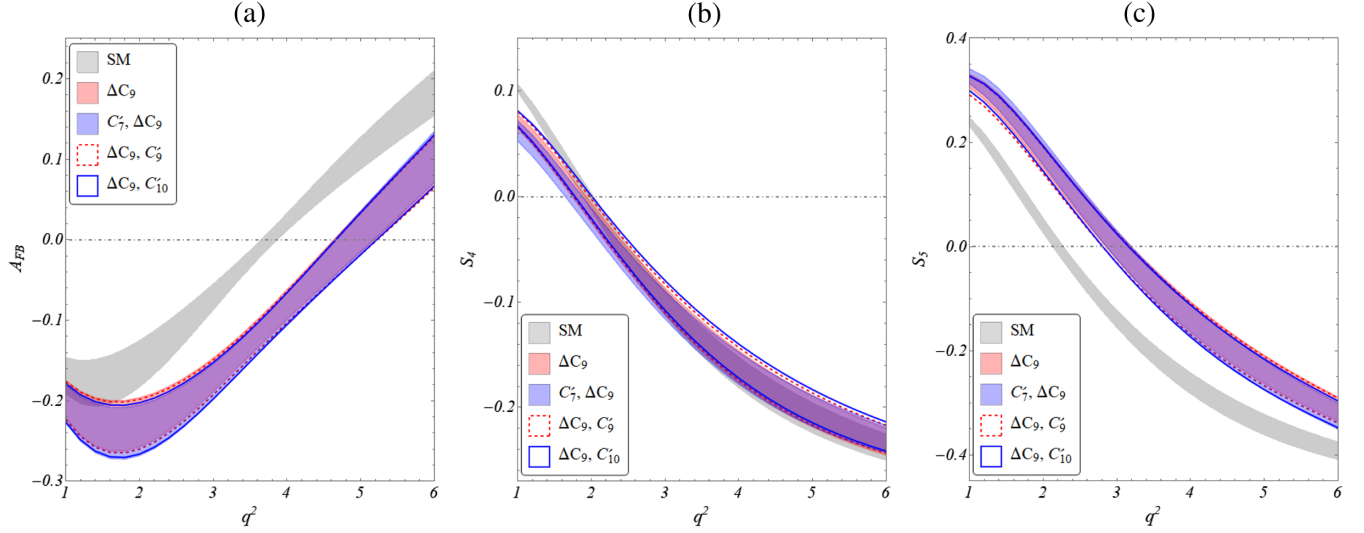
FIG. 8. q^2 distributions of a few angular observables.

We should also mention here that the scenario $\Delta C_9 = -\Delta C_{10}$ that arises, for example, in some leptoquark models (among other model-dependent origins) does not pass the criteria of $\Delta AIC_c \leq 4$. We hence do not display or discuss this scenario.

VI. SUMMARY

In this article, we have analyzed the semileptonic $b \to s\ell\ell$ decays in a model independent framework with the relevant dimension six effective operators invariant under the strong and electromagnetic gauge groups. Our chosen set of operators does not include the four quark operators, chromomagnetic operators, and tensor operators. Different possible combinations of all the effective operators have been considered, and following the statistical tools like cross-validation and the small-sample-corrected Akaike information criterion (AIC_c), we have found out the combinations which best explain the available data. We have provided separate analyses for the data on angular observables obtained from an unbinned maximum likelihood fit and that due to the principal moments of the angular distribution in $B \to K^*\mu^+\mu^-$ decay.

Among all the possible combinations, a relatively small number of one, two, and three-operator scenarios satisfy the criterion of a selected "best" model. All the selected scenarios contain a left-handed quark current with vector muon coupling as an operator ($\mathcal{O}_9$). This is also the only one-operator scenario that survives the exclusion test in our search for the best model(s). We have noted differences between the selected models in the analysis with angular data from likelihood fit and those from the principal moments analysis. The $R_{K^{(*)}}$, along with the angular observables associated with $B \to K^*\mu^+\mu^-$ decays, have played an important role in this selection. In the analysis

with the new data on $R_{K^{(*)}}$, the scenarios with three, four and five operators are selected. This could be due to the fact that the tension between the updated measured values of R_K and R_{K^*} and their respective SM predictions have reduced in comparison to that for their old experimental measurements. Therefore, in order to explain all the data simultaneously, simultaneous contributions from different operators are required. We have noticed changes in the selected scenarios when we drop $R_{K^{(*)}}$ from the list of inputs. We have performed the analysis with and without the 2019 updates on $R_{K^{(*)}}$ from Belle and have compared them. We have noticed changes in the allowed parameter spaces for the Wilson coefficients of the selected scenarios.

We have compared our fit results for the angular observables with the corresponding SM predictions and the measured values in different bins for a few selected models. While our fit results are fully consistent with both the measured values and the respective SM predictions in the analysis of likelihood data, in the analysis of moment data, there are discrepancies between our fitted results and the respective SM predictions and the measured values in some bins. For some of the selected scenarios, we have studied the correlations between different observables, which show that the operator $\mathcal{O}_9$ and the combination of $\mathcal{O}_9$ and $\mathcal{O}_7'$ (flavor changing electromagnetic dipole operator) cannot explain all the available data on $R_{K^{(*)}}$ simultaneously. In particular, they have difficulty in explaining the observed results of R_{K^*} in the low q^2 bin ($q^2 \in [0.045, 1.1]$ GeV^2).

We have studied the NP effects in $R_{K^{(*)}}$ only, and noticed that the operator with a left-handed quark current with an axial-vector muon coupling ($\mathcal{O}_{10}$) is the only one-operator scenario that can explain the data. Also, the parameter space for the corresponding Wilson-coefficient ΔC_{10}, allowed by $R_{K^{(*)}}$, is tightly constrained by the measured

values of $\mathrm{Br}(B_s \to \mu^+\mu^-)$. However, there are a few two-operator scenarios which have the potential to explain the current observation. Those operators are obtained from possible combinations of $\mathcal{O}_9$, $\mathcal{O}_{10}$ and operators like right-handed quark current with vector or axial-vector muon couplings ($\mathcal{O}'_9$, $\mathcal{O}'_{10}$). In the two-operator scenarios, the allowed parameter spaces for ΔC_{10} or/and C'_{10} can comfortably explain the observed data on $\mathrm{Br}(B_s \to \mu^+\mu^-)$.

ACKNOWLEDGMENTS

Note added.—The complete list of all models used in this analysis is too long to include in this draft. We have added an ancillary file named "`models.json`" along with this draft (to be found within the arXiv source file). This file contains all combinations of WCs, relating them with their corresponding indices in our analysis.

APPENDIX A: BEST FIT VALUES OF THE NEW WCs IN THE ANALYSIS WITH OLD DATASET

The selected models in the analysis with old datasets are shown in Fig. 4. The best fit values of the corresponding WCs along with their respective errors are given in Tables IV and V, respectively.

TABLE IV. Fit-qualities, model selection criteria, parameter estimates and effects on radiative decays for the best selected models with the old dataset, with the moments estimate of the angular observables. Selected models are obtained from Fig. 4(a). Last four columns showcase the deviations (in units of σ) between the experimental value of the radiative decays and the corresponding value obtained with the fit results.

Model index	$\chi^2_{\mathrm{Min}}/\mathrm{DOF}$	p-val (%)	$\omega^{\Delta \mathrm{AIC}_c}$ (%)	$\mathrm{MSE}_{X\text{-val}}$	Parameter values	Deviation in σ			
						$B \to X_s\gamma$	$B^+ \to K^*\gamma$	$\Delta B^0 \to K^*\gamma$	$\Delta B_s \to \phi\gamma$
2	245.67/254	63.5	5.	0.918	$\Delta C_9 \to -1.26 \pm 0.14$	–	–	–	–
10	244.92/253	63.1	2.6	0.916	$C'_7 \to 0.013 \pm 0.015$ $\Delta C_9 \to -1.3 \pm 0.15$	0.32	−0.87	−1.06	1.22
19	245.42/253	62.2	2.	0.926	$\Delta C_9 \to -1.22 \pm 0.16$ $\Delta C_{10} \to 0.061 \pm 0.123$	–	–	–	–
21	245.48/253	62.1	2.	0.923	$\Delta C_9 \to -1.27 \pm 0.15$ $C_S \to -0.021 \pm 0.026$	–	–	–	–
22	245.51/253	62.	1.9	0.923	$\Delta C_9 \to -1.27 \pm 0.15$ $C'_S \to 0.02 \pm 0.026$	–	–	–	–
18	245.55/253	62.	1.9	0.915	$\Delta C_9 \to -1.25 \pm 0.14$ $C'_9 \to 0.067 \pm 0.195$	–	–	–	–
20	245.59/253	61.9	1.9	0.92	$\Delta C_9 \to -1.26 \pm 0.14$ $C'_{10} \to -0.03 \pm 0.109$	–	–	–	–

TABLE V. Same as Table IV, but with the likelihood estimate of the angular observables. Selected models are obtained from Fig. 4(b).

| Model index	$\chi^2_{\mathrm{Min}}/\mathrm{DOF}$	p-val (%)	$\omega^{\Delta \mathrm{AIC}_c}$ (%)	$\mathrm{MSE}_{X\text{-val}}$	Parameter values	Deviation in σ			
						$B \to X_s\gamma$	$B^+ \to K^*\gamma$	$\Delta B^0 \to K^*\gamma$	$\Delta B_s \to \phi\gamma$
10	213.78/209	39.6	5.3	0.973	$C'_7 \to 0.028 \pm 0.015$ $\Delta C_9 \to -1.37 \pm 0.14$	0.37	−0.85	−1.04	1.24
2	217.19/210	35.2	2.7	0.989	$\Delta C_9 \to -1.28 \pm 0.13$	–	–	–	–
49	213.2/208	38.8	2.5	0.981	$C'_7 \to 0.029 \pm 0.015$ $\Delta C_9 \to -1.4 \pm 0.14$ $C_S \to -0.028 \pm 0.019$	0.38	−0.85	−1.04	1.25

(Table continued)

TABLE V. *(Continued)*

Model index	$\chi^2_{\text{Min}}/\text{DOF}$	p-val (%)	$\omega^{\Delta\text{AIC}_c}$ (%)	$\text{MSE}_{X\text{-val}}$	Parameter values	Deviation in σ			
						$B \to X_s\gamma$	$B^+ \to K^*\gamma$	$\Delta B^0 \to K^*\gamma$	$\Delta B_s \to \phi\gamma$
50	213.23/208	38.7	2.5	0.981	$C'_7 \to 0.029 \pm 0.015$ $\Delta C_9 \to -1.4 \pm 0.14$ $C'_S \to 0.028 \pm 0.019$	0.38	-0.85	-1.04	1.25
47	213.65/208	37.9	2.	0.976	$C'_7 \to 0.029 \pm 0.015$ $\Delta C_9 \to -1.39 \pm 0.15$ $\Delta C_{10} \to -0.042 \pm 0.117$	0.38	-0.85	-1.04	1.25

APPENDIX B: RESULTS FOR MODEL-SELECTION WITH DATASETS OTHER THAN NEW-MOMENTS AND NEW-LIKELIHOOD

A comparison of the selected models in Figs. 3 and 9 is helpful in understanding the impact of the new $R_{K^{(*)}}$ measurements by LHCb and Belle on the process of model selection. With more precise measurements on R_K and R_{K^*}, the consistency between the data and the corresponding SM measurements may increase further. Therefore, it is important to gain insights on the probable NP effects in the angular observables. In Fig. 9, we provide the results for the model selection corresponding to the "new-moments" and the "new-likelihood" datasets after dropping the LFUV observables R_K and R_{K^*}. The selected scenarios can be compared with the one given in Fig. 3. We note that in both the likelihood and moments data under the given selection setup, the number of selected scenarios reduce after we drop $R(K)$ and $R(K^*)$ from the inputs. A comparison of between Figs. 9(a) and 3(a) indicates that the three-operator scenarios become less favorable. Similarly, from Figs. 9(b) and 3(b) we see that the four and five-operators scenarios are less favorable in the analysis without $R(K)$ and $R(K^*)$. As explained in the main text, the explanation of new data

prefers NP scenarios with more than two operators like three, four, or five-operator scenarios. In particular, we have noted that the fit qualities improve once we drop these LFUV observables from the fits in general, which is on par with our expectations. For comparison with the new dataset, Fig. 10 lists the angular observables from SM, experiment, and our fit results.

Figure 11 depicts the allowed parameter spaces of the most commonly occurring one and two parameter scenarios selected from different types of fits and datasets. For one operator scenario ($\mathcal{O}_9$), the allowed parameter space of the corresponding WC ΔC_9, is shown in Fig. 11(a). For the two operator scenarios, Figs. 11(b), 11(c), and 11(d) shows the correlations between the WCs. We note that the allowed values of C'_7 and ΔC_9 have reasonably small ranges and they are negatively correlated. The corresponding value is -0.316. Large (negative) values of ΔC_9 prefers large positive values of C'_7. The other two plots show the allowed parameter spaces and the correlations of ΔC_9 with C'_9 with C'_{10}, respectively. The fitted values of C'_9 and C'_{10} have large errors. Guessing the exact correlations between them from the figures alone is therefore difficult. However, one can see that in the analysis with new data, the value of the
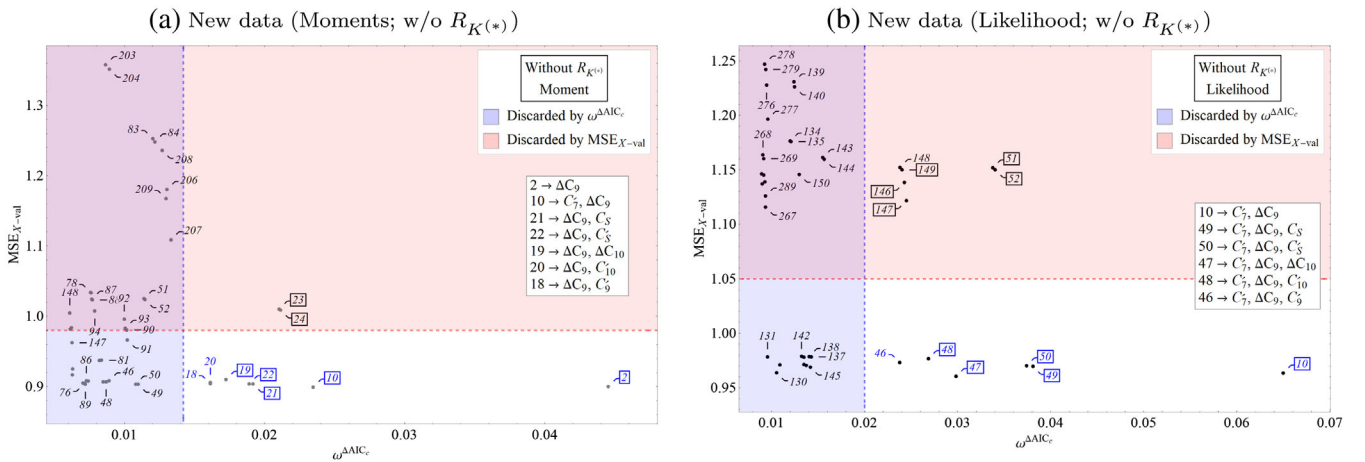
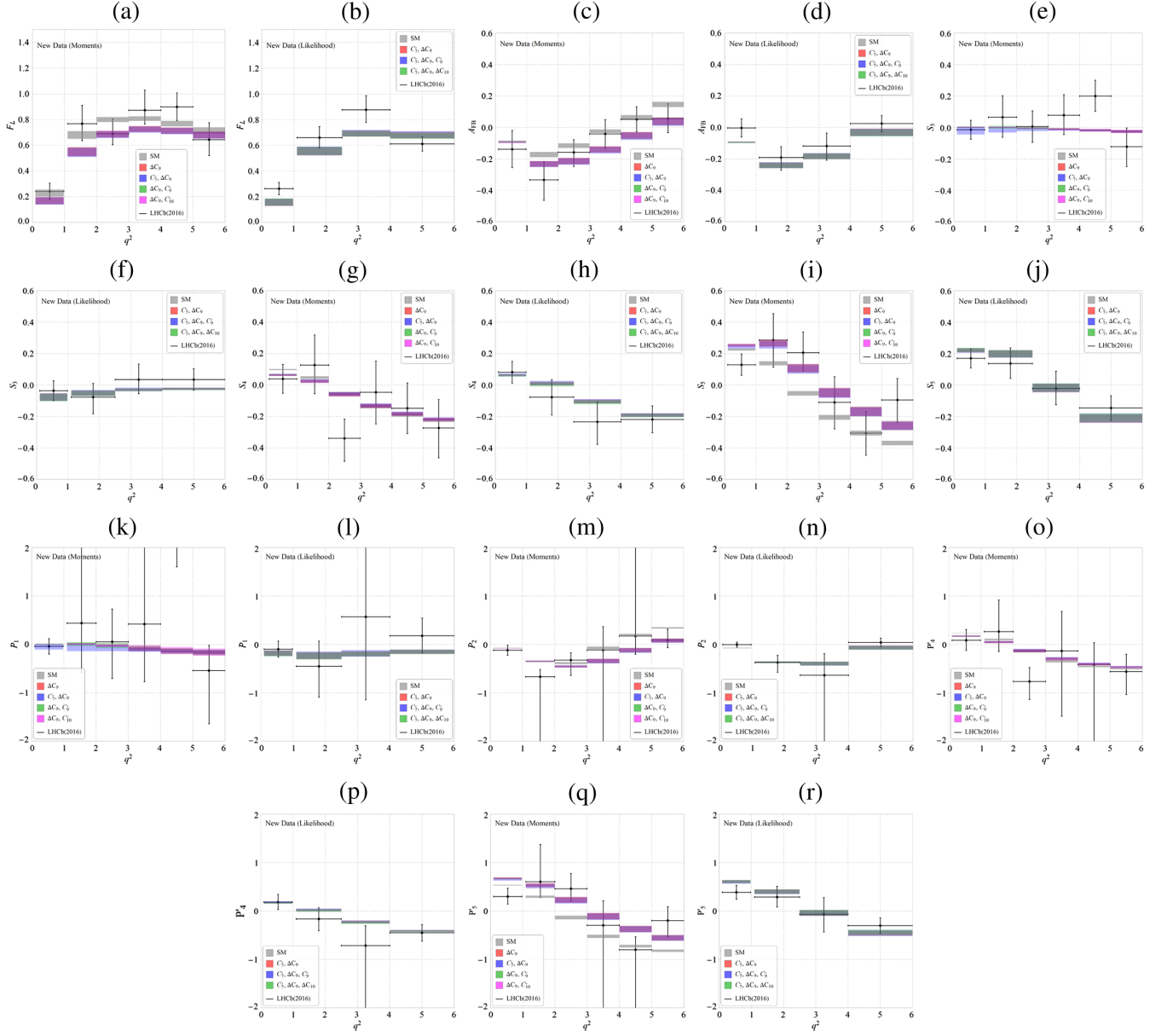FIG. 9. Results similar to Fig. 3 but with R_K and R_{K^*} dropped.

FIG. 10. Comparison of the *CP*-averaged angular observables in different bins which are obtained in experiment, SM and from our fit results after dropping $R_{K^{(*)}}$.
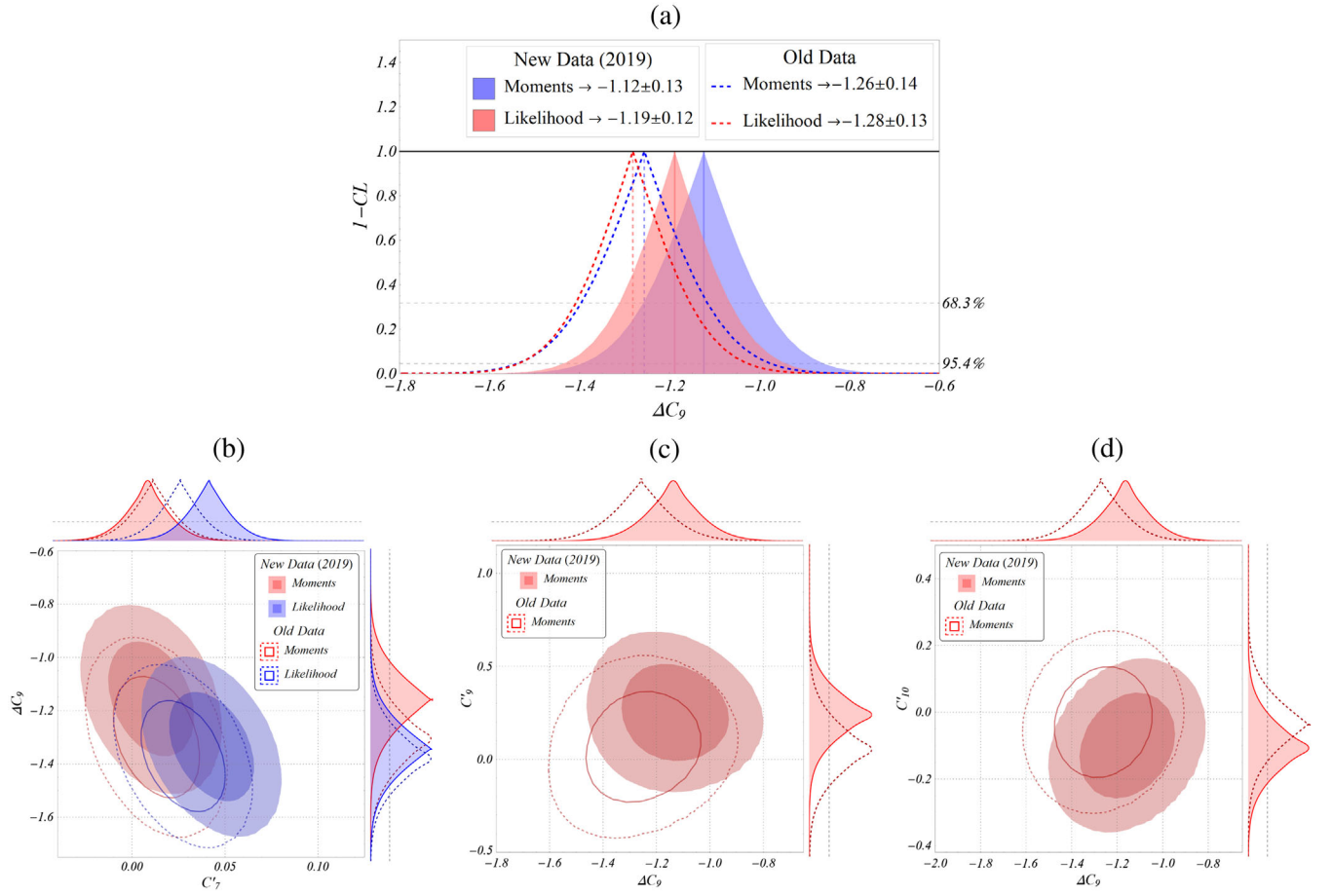
FIG. 11. Allowed parameter space for ΔC_9 in one-operator scenario and allowed NP parameter spaces and their respective correlation for some selected two-operator scenarios.
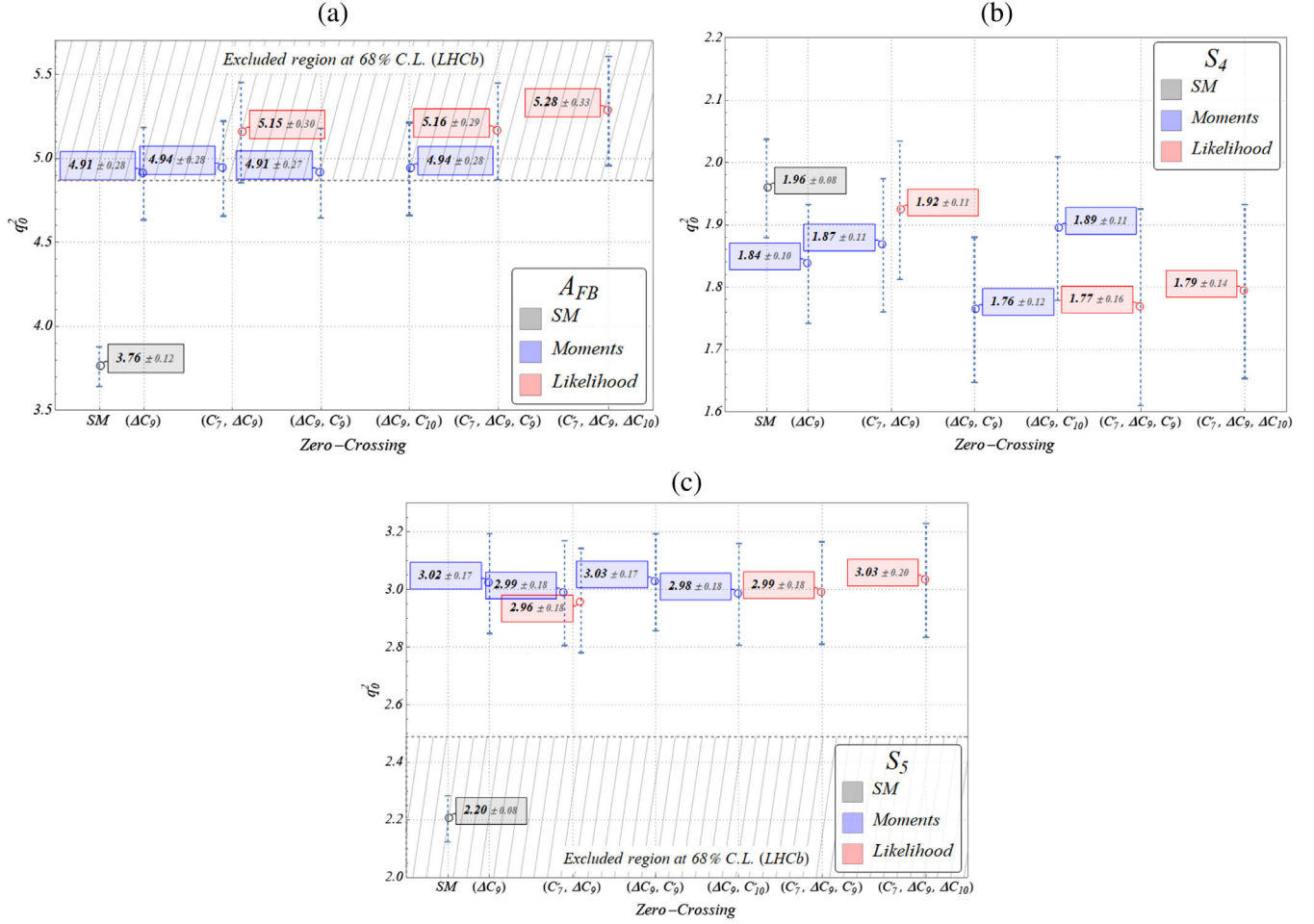
FIG. 12. Zero crossing values in the q^2 distributions of the angular observables shown in Fig. 8. q_0^2 represents the value of q^2 at the zero crossing.

correlation between ΔC_9 and C'_{10} (+0.24) is greater than that between ΔC_9 and C'_9 which is -0.11.

In Fig. 12 we compare the values of q^2 at the zero crossing (q_0^2) between SM, our selected models and the

measured values for the above mentioned observables. We note that in our selected models, the q_0^2 for all these three observables are in good agreement with the corresponding measured values.

[1] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 02 (2016) 104.

[2] M. Aaboud *et al.* (ATLAS Collaboration), J. High Energy Phys. 10 (2018) 047.

[3] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, J. High Energy Phys. 01 (2013) 048.

[4] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, J. High Energy Phys. 05 (2013) 137.

[5] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, Phys. Rev. Lett. **112**, 212003 (2014).

[6] A. Bharucha, D. M. Straub, and R. Zwicky, J. High Energy Phys. 08 (2016) 098.

[7] M. Ciuchini, M. Fedele, E. Franco, S. Mishima, A. Paul, L. Silvestrini, and M. Valli, J. High Energy Phys. 06 (2016) 116.

[8] G. Hiller and F. Kruger, Phys. Rev. D **69**, 074020 (2004).

[9] M. Bordone, G. Isidori, and A. Pattori, Eur. Phys. J. C **76**, 440 (2016).

[10] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 08 (2017) 055.

[11] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **122**, 191801 (2019).

[12] A. Abdesselam *et al.* (Belle Collaboration), arXiv:1904 .02440.

[13] A. Abdesselam *et al.*, arXiv:1908.01848.

[14] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, and D. M. Straub, arXiv:1903.10434.

[15] A. K. Alok, A. Dighe, S. Gangal, and D. Kumar, J. High Energy Phys. 06 (2019) 089.

[16] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto, J. High Energy Phys. 01 (2018) 093.

[17] A. Arbey, T. Hurth, F. Mahmoudi, and S. Neshatpour, Phys. Rev. D **98,** 095027 (2018).

[18] S. Descotes-Genon, J. Matias, and J. Virto, Phys. Rev. D **88,** 074002 (2013).

[19] W. Altmannshofer and D. M. Straub, Eur. Phys. J. C **73,** 2646 (2013).

[20] F. Beaujean, C. Bobeth, and D. van Dyk, Eur. Phys. J. C **74,** 2897 (2014); **74,** 3179(E) (2014).

[21] T. Hurth and F. Mahmoudi, J. High Energy Phys. 04 (2014) 097.

[22] W. Altmannshofer and D. M. Straub, Eur. Phys. J. C **75,** 382 (2015).

[23] W. Altmannshofer, P. Ball, A. Bharucha, A. J. Buras, D. M. Straub, and M. Wick, J. High Energy Phys. 01 (2009) 019.

[24] U. Egede, T. Hurth, J. Matias, M. Ramon, and W. Reece, J. High Energy Phys. 10 (2010) 056.

[25] D. Das and R. Sinha, Phys. Rev. D **86,** 056006 (2012).

[26] T. Hurth, F. Mahmoudi, and S. Neshatpour, J. High Energy Phys. 12 (2014) 053.

[27] M. Ciuchini, A. M. Coutinho, M. Fedele, E. Franco, A. Paul, L. Silvestrini, and M. Valli, Eur. Phys. J. C **79,** 719 (2019).

[28] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker, J. High Energy Phys. 01 (2012) 107.

[29] L.-S. Geng, B. Grinstein, S. Jäger, J. Martin Camalich, X.-L. Ren, and R.-X. Shi, Phys. Rev. D **96,** 093006 (2017).

[30] M. Ciuchini, A. M. Coutinho, M. Fedele, E. Franco, A. Paul, L. Silvestrini, and M. Valli, Eur. Phys. J. C **77,** 688 (2017).

[31] K. Kowalska, D. Kumar, and E. M. Sessolo, Eur. Phys. J. C **79,** 840 (2019).

[32] G. E. P. Box, J. Am. Stat. Assoc. **71,** 791 (1976).

[33] G. E. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters: Design, Discovery and Innovation* (Wiley-Interscience, Hoboken, 2005), Vol. 639.

[34] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, New York, 2002).

[35] S. Geisser and W. F. Eddy, J. Am. Stat. Assoc. **74,** 153 (1979).

[36] R. Andrae, T. Schulze-Hartung, and P. Melchior, arXiv:1012.3754.

[37] H. Akaike, *Selected Papers of Hirotugu Akaike* (Springer, New York, 1998).

[38] S. Arlot and A. Celisse, Stat. Surveys **4,** 40 (2010).

[39] S. Bhattacharya, S. Nandi, and S. K. Patra, Phys. Rev. D **95,** 075012 (2017).

[40] S. Bhattacharya, S. Nandi, and S. Kumar Patra, Eur. Phys. J. C **79,** 268 (2019).

[41] A. Gelman, J. Hwang, and A. Vehtari, arXiv:1307.5928.

[42] T. Ando, Biometrika **94,** 443 (2007).

[43] G. Claeskens and N. L. Hjort, The bayesian information criterion, in *Model Selection and Model Averaging, Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge University Press, Cambridge, England, 2008), pp. 70–98.

[44] C. Collaboration *et al.*, Cdf public note 10894 (2012).

[45] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 11 (2016) 047; 04 (2017) 142(E).

[46] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 06 (2014) 133.

[47] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. D **98,** 112011 (2018).

[48] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 09 (2015) 179.

[49] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **113,** 151601 (2014).

[50] Hflav $b_s \to \mu\mu$ average, https://hflav-eos.web.cern.ch/hflav-eos/rare/April2019/BS/OUTPUT/HTML/bs_table1.html.

[51] J. Charles, S. Descotes-Genon, Z. Ligeti, S. Monteil, M. Papucci, and K. Trabelsi, Phys. Rev. D **89,** 033016 (2014).

[52] S. Aoki *et al.* (Flavour Lattice Averaging Group), Eur. Phys. J. C **80,** 113 (2020).

[53] F. Beaujean, M. Chrzaszcz, N. Serra, and D. van Dyk, Phys. Rev. D **91,** 114012 (2015).

[54] J. Gratrex, M. Hopfer, and R. Zwicky, Phys. Rev. D **93,** 054008 (2016).

[55] B. Efron, Ann. Stat. **7,** 1 (1979).

[56] N. Gubernari, A. Kokulu, and D. van Dyk, J. High Energy Phys. 01 (2019) 150.

[57] C. Bobeth, G. Hiller, and G. Piranishvili, J. High Energy Phys. 12 (2007) 040.

[58] C. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu (HPQCD Collaboration), Phys. Rev. D **88,** 054509 (2013); **88,** 079901(E) (2013).

[59] P. Ball and R. Zwicky, Phys. Rev. D **71,** 014015 (2005).

[60] M. Bartsch, M. Beylich, G. Buchalla, and D. N. Gao, J. High Energy Phys. 11 (2009) 011.

[61] A. J. Buras, J. Girrbach-Noe, C. Niehoff, and D. M. Straub, J. High Energy Phys. 02 (2015) 184.

[62] A. Paul and D. M. Straub, J. High Energy Phys. 04 (2017) 027.

[63] M. Misiak *et al.*, Phys. Rev. Lett. **114,** 221801 (2015).

[64] M. Misiak and M. Steinhauser, Eur. Phys. J. C **77,** 201 (2017).

[65] Y. Amhis *et al.* (Heavy Flavor Averaging Group (HFAG) Collaboration), arXiv:1412.7515.

[66] R. Aaij *et al.* (LHCb Collaboration), Nucl. Phys. **B867,** 1 (2013).

[67] D. Dutta *et al.* (Belle Collaboration), Phys. Rev. D **91,** 011101 (2015).

[68] R. Alonso, B. Grinstein, and J. Martin Camalich, Phys. Rev. Lett. **113,** 241802 (2014).

[69] S. K. Patra, OptEx (2019), still under development, updates will be available in the url in future.

[70] R. Storn and K. Price, J. Global Optim. **11,** 341 (1997).

[71] E. Gross, *CERN Yellow Reports: School Proceedings* (CERN, Geneva, 2017), Vol. 4, p. 165.

[72] L. Demortier and L. Lyons (CDF Collaboration), CDF/ANAL/PUBLIC/5776 B773 (2008).

[73] C. Harald, Scand. Actuarial J. **1928,** 13 (1928).

[74] F. James and M. Roos, Comput. Phys. Commun. **10,** 343 (1975).

[75] R. Aaij *et al.* (LHCb Collaboration), J. High Energy Phys. 12 (2016) 087.

[76] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, London, 1970).

[77] L. Breiman, J. Am. Stat. Assoc. **87**, 738 (1992).

[78] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, Anal. Chim. Acta **760**, 25 (2013).

[79] G. Varoquaux, arXiv:1706.07581.

[80] N. Sugiura, Commun. Stat., Theory Methods **7**, 13 (1978).

[81] R. Shibata, Biometrika **68**, 45 (1981).