

# Dark Energy Survey Year 3 Results: Constraints on extensions to $\Lambda$ CDM with weak lensing and galaxy clustering

(DES Collaboration)

T. M. C. Abbott,<sup>1</sup> M. Aguena,<sup>2</sup> A. Alarcon,<sup>3</sup> O. Alves,<sup>4,2</sup> A. Amon,<sup>5,6</sup> F. Andrade-Oliveira,<sup>4</sup> J. Annis,<sup>7</sup> S. Avila,<sup>8</sup> D. Bacon,<sup>9</sup> E. Baxter,<sup>10</sup> K. Bechtol,<sup>11</sup> M. R. Becker,<sup>3</sup> G. M. Bernstein,<sup>12</sup> S. Birrer,<sup>13,14</sup> J. Blazek,<sup>15</sup> S. Bocquet,<sup>16</sup> A. Brandao-Souza,<sup>17,2</sup> S. L. Bridle,<sup>18</sup> D. Brooks,<sup>19</sup> D. L. Burke,<sup>13,14</sup> H. Camacho,<sup>20,2</sup> A. Campos,<sup>21</sup> A. Carnero Rosell,<sup>22,2,23</sup> M. Carrasco Kind,<sup>24,25</sup> J. Carretero,<sup>26</sup> F. J. Castander,<sup>27,28</sup> R. Cawthon,<sup>29</sup> C. Chang,<sup>30,31</sup> A. Chen,<sup>4,32</sup> R. Chen,<sup>33</sup> A. Choi,<sup>34</sup> C. Conselice,<sup>18,35</sup> J. Cordero,<sup>18</sup> M. Costanzi,<sup>36,37,38</sup> M. Crocce,<sup>27,28</sup> L. N. da Costa,<sup>2</sup> M. E. S. Pereira,<sup>39</sup> C. Davis,<sup>13</sup> T. M. Davis,<sup>40</sup> J. DeRose,<sup>41</sup> S. Desai,<sup>42</sup> E. Di Valentino,<sup>18</sup> H. T. Diehl,<sup>7</sup> S. Dodelson,<sup>21,43</sup> P. Doel,<sup>19</sup> C. Dux,<sup>12</sup> A. Drlica-Wagner,<sup>30,7,31</sup> K. Eckert,<sup>12</sup> T. F. Eifler,<sup>44,45</sup> F. Elsner,<sup>19</sup> J. Elvin-Poole,<sup>46,47</sup> S. Everett,<sup>45</sup> X. Fang,<sup>48,44</sup> A. Farahi,<sup>49</sup> I. Ferrero,<sup>50</sup> A. Ferté,<sup>45</sup> B. Flaugher,<sup>7</sup> P. Fosalba,<sup>27,28</sup> D. Friedel,<sup>24</sup> O. Friedrich,<sup>6</sup> J. Frieman,<sup>7,31</sup> J. García-Bellido,<sup>8</sup> M. Gatti,<sup>12</sup> L. Giani,<sup>40</sup> T. Giannantonio,<sup>5,6</sup> G. Giannini,<sup>26</sup> D. Gruen,<sup>16</sup> R. A. Gruendl,<sup>24,25</sup> J. Gschwend,<sup>2,51</sup> G. Gutierrez,<sup>7</sup> N. Hamaus,<sup>52</sup> I. Harrison,<sup>53,18,54</sup> W. G. Hartley,<sup>55</sup> K. Herner,<sup>7</sup> S. R. Hinton,<sup>56</sup> D. L. Hollowood,<sup>57</sup> K. Honscheid,<sup>46,47</sup> H. Huang,<sup>44,58</sup> E. M. Huff,<sup>45</sup> D. Huterer,<sup>4</sup> B. Jain,<sup>12</sup> D. J. James,<sup>59</sup> M. Jarvis,<sup>12</sup> N. Jeffrey,<sup>19</sup> T. Jeltima,<sup>57</sup> A. Kovacs,<sup>22,23</sup> E. Krause,<sup>44</sup> K. Kuehn,<sup>60,61</sup> N. Kuropatkin,<sup>7</sup> O. Lahav,<sup>19</sup> S. Lee,<sup>33</sup> P.-F. Leget,<sup>13</sup> P. Lemos,<sup>19,62</sup> C. D. Leonard,<sup>63</sup> A. R. Liddle,<sup>64</sup> M. Lima,<sup>65,2</sup> H. Lin,<sup>7</sup> N. MacCrann,<sup>66</sup> J. L. Marshall,<sup>67</sup> J. McCullough,<sup>13</sup> J. Mena-Fernández,<sup>68</sup> F. Menanteau,<sup>24,25</sup> R. Miquel,<sup>69,26</sup> V. Miranda,<sup>70</sup> J. J. Mohr,<sup>71,16</sup> J. Muir,<sup>72</sup> J. Myles,<sup>73,13,14</sup> S. Nadathur,<sup>9</sup> A. Navarro-Alsina,<sup>17</sup> R. C. Nichol,<sup>9</sup> R. L. C. Ogando,<sup>51</sup> Y. Omori,<sup>30,73,31,13</sup> A. Palmese,<sup>48</sup> S. Pandey,<sup>12</sup> Y. Park,<sup>32</sup> M. Paterno,<sup>7</sup> F. Paz-Chinchón,<sup>24,5</sup> W. J. Percival,<sup>74,72</sup> A. Pieres,<sup>2,51</sup> A. A. Plazas Malagón,<sup>75</sup> A. Porredon,<sup>46,47</sup> J. Prat,<sup>30,31</sup> M. Raveri,<sup>12</sup> M. Rodríguez-Monroy,<sup>68</sup> P. Rogozinski,<sup>58</sup> R. P. Rollins,<sup>18</sup> A. K. Romer,<sup>62</sup> A. Roodman,<sup>13,14</sup> R. Rosenfeld,<sup>76,2</sup> A. J. Ross,<sup>46</sup> E. S. Rykoff,<sup>13,14</sup> S. Samuroff,<sup>21</sup> C. Sánchez,<sup>12</sup> E. Sanchez,<sup>68</sup> J. Sanchez,<sup>7</sup> D. Sanchez Cid,<sup>68</sup> V. Scarpine,<sup>7</sup> D. Scolnic,<sup>33</sup> L. F. Secco,<sup>31</sup> I. Sevilla-Noarbe,<sup>68</sup> E. Sheldon,<sup>77</sup> T. Shin,<sup>78</sup> M. Smith,<sup>79</sup> M. Soares-Santos,<sup>4</sup> E. Suchyta,<sup>80</sup> M. Tabbutt,<sup>11</sup> G. Tarle,<sup>4</sup> D. Thomas,<sup>9</sup> C. To,<sup>46</sup> A. Troja,<sup>76,2</sup> M. A. Troxel,<sup>33</sup> I. Tutusaus,<sup>81,27,28</sup> T. N. Varga,<sup>82,71,52</sup> M. Vincenzi,<sup>9,79</sup> A. R. Walker,<sup>1</sup> N. Weaverdyck,<sup>4,41</sup> R. H. Wechsler,<sup>73,13,14</sup> J. Weller,<sup>71,52</sup> B. Yanny,<sup>7</sup> B. Yin,<sup>21</sup> Y. Zhang,<sup>67</sup> and J. Zuntz<sup>83</sup>

(DES Collaboration)

<sup>1</sup>Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile

<sup>2</sup>Laboratório Interinstitucional de e-Astronomia - LIneA,

Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil

<sup>3</sup>Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

<sup>4</sup>Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>5</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>6</sup>Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

<sup>7</sup>Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA

<sup>8</sup>Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, 28049 Madrid, Spain

<sup>9</sup>Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK

<sup>10</sup>Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

<sup>11</sup>Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390

<sup>12</sup>Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>13</sup>Kavli Institute for Particle Astrophysics & Cosmology,

P. O. Box 2450, Stanford University, Stanford, CA 94305, USA

<sup>14</sup>SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>15</sup>Department of Physics, Northeastern University, Boston, MA 02115, USA

<sup>16</sup>University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany

<sup>17</sup>Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil

<sup>18</sup>Jodrell Bank Center for Astrophysics, School of Physics and Astronomy,

University of Manchester, Oxford Road, Manchester, M13 9PL, UK

<sup>19</sup>Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

<sup>20</sup>Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil

<sup>21</sup>Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA

<sup>22</sup>Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

<sup>23</sup>Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain

<sup>24</sup>Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA

<sup>25</sup>Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA

<sup>26</sup>Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain

<sup>27</sup>Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain

- <sup>28</sup>*Institute of Space Sciences (ICE, CSIC), Campus UAB,  
Carrer de Can Magrans, s/n, 08193 Barcelona, Spain*
- <sup>29</sup>*Physics Department, William Jewell College, Liberty, MO, 64068*
- <sup>30</sup>*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
- <sup>31</sup>*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
- <sup>32</sup>*Kavli Institute for the Physics and Mathematics of the Universe (WPI),  
UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
- <sup>33</sup>*Department of Physics, Duke University Durham, NC 27708, USA*
- <sup>34</sup>*California Institute of Technology, 1200 East California Blvd, MC 249-17, Pasadena, CA 91125, USA*
- <sup>35</sup>*University of Nottingham, School of Physics and Astronomy, Nottingham NG7 2RD, UK*
- <sup>36</sup>*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*
- <sup>37</sup>*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
- <sup>38</sup>*Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy*
- <sup>39</sup>*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany*
- <sup>40</sup>*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
- <sup>41</sup>*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
- <sup>42</sup>*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
- <sup>43</sup>*NSF AI Planning Institute for Physics of the Future,  
Carnegie Mellon University, Pittsburgh, PA 15213, USA*
- <sup>44</sup>*Department of Astronomy/Steward Observatory, University of Arizona,  
933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
- <sup>45</sup>*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
- <sup>46</sup>*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
- <sup>47</sup>*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
- <sup>48</sup>*Department of Astronomy, University of California, Berkeley, 501 Campbell Hall, Berkeley, CA 94720, USA*
- <sup>49</sup>*Departments of Statistics and Data Science, University of Texas at Austin, Austin, TX 78757, USA*
- <sup>50</sup>*Institute of Theoretical Astrophysics, University of Oslo. P.O. Box 1029 Blindern, NO-0315 Oslo, Norway*
- <sup>51</sup>*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*
- <sup>52</sup>*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany*
- <sup>53</sup>*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
- <sup>54</sup>*School of Physics and Astronomy, Cardiff University, CF24 3AA, UK*
- <sup>55</sup>*Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland*
- <sup>56</sup>*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072*
- <sup>57</sup>*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
- <sup>58</sup>*Department of Physics, University of Arizona, Tucson, AZ 85721, USA*
- <sup>59</sup>*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
- <sup>60</sup>*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
- <sup>61</sup>*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
- <sup>62</sup>*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton, BN1 9QH, UK*
- <sup>63</sup>*School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK*
- <sup>64</sup>*Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, 1769-016 Lisboa, Portugal*
- <sup>65</sup>*Departamento de Física Matemática, Instituto de Física,  
Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil*
- <sup>66</sup>*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
- <sup>67</sup>*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy,  
and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
- <sup>68</sup>*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
- <sup>69</sup>*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
- <sup>70</sup>*C. N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, NY 11794*
- <sup>71</sup>*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*
- <sup>72</sup>*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
- <sup>73</sup>*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
- <sup>74</sup>*Department of Physics and Astronomy, University of Waterloo,  
200 University Ave W, Waterloo, ON N2L 3G1, Canada*
- <sup>75</sup>*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
- <sup>76</sup>*ICTP South American Institute for Fundamental Research  
Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
- <sup>77</sup>*Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*
- <sup>78</sup>*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
- <sup>79</sup>*School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK*
- <sup>80</sup>*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831*
- <sup>81</sup>*Département de Physique Théorique and Center for Astroparticle Physics,  
Université de Genève, 24 quai Ernest Ansermet, CH-1211 Geneva, Switzerland*
- <sup>82</sup>*Excellence Cluster Origins, Boltzmannstr. 2, 85748 Garching, Germany*

We constrain six possible extensions to the  $\Lambda$ CDM model using measurements from the Dark Energy Survey’s first three years of observations, alone and in combination with external cosmological probes. The DES data are the two-point correlation functions of weak gravitational lensing, galaxy clustering, and their cross-correlation. We use simulated data vectors and blind analyses of real data to validate the robustness of our results to astrophysical and modeling systematic errors. In many cases, constraining power is limited by the absence of theoretical predictions beyond the linear regime that are reliable at our required precision. The  $\Lambda$ CDM extensions are: dark energy with a time-dependent equation of state, non-zero spatial curvature, additional relativistic degrees of freedom, sterile neutrinos with eV-scale mass, modifications of gravitational physics, and a binned  $\sigma_8(z)$  model which serves as a phenomenological probe of structure growth. For the time-varying dark energy equation of state evaluated at the pivot redshift we find  $(w_p, w_a) = (-0.99_{-0.17}^{+0.28}, -0.9 \pm 1.2)$  at 68% confidence with  $z_p = 0.24$  from the DES measurements alone, and  $(w_p, w_a) = (-1.03_{-0.03}^{+0.04}, -0.4_{-0.3}^{+0.4})$  with  $z_p = 0.21$  for the combination of all data considered. Curvature constraints of  $\Omega_k = 0.0009 \pm 0.0017$  and effective relativistic species  $N_{\text{eff}} = 3.10_{-0.16}^{+0.15}$  are dominated by external data, though adding DES information to external low redshift probes tightens the  $\Omega_k$  constraints that can be made without CMB observables by 20%. For massive sterile neutrinos, DES combined with external data improves the upper bound on the mass  $m_{\text{eff}}$  by a factor of three compared to previous analyses, giving 95% limits of  $(\Delta N_{\text{eff}}, m_{\text{eff}}) \leq (0.28, 0.20 \text{ eV})$  when using priors matching a comparable *Planck* analysis. For modified gravity, we constrain changes to the lensing and Poisson equations controlled by functions  $\Sigma(k, z) = \Sigma_0 \Omega_\Lambda(z)/\Omega_{\Lambda,0}$  and  $\mu(k, z) = \mu_0 \Omega_\Lambda(z)/\Omega_{\Lambda,0}$  respectively to  $\Sigma_0 = 0.6_{-0.5}^{+0.4}$  from DES alone and  $(\Sigma_0, \mu_0) = (0.04 \pm 0.05, 0.08_{-0.19}^{+0.21})$  for the combination of all data, both at 68% confidence. Overall, we find no significant evidence for physics beyond  $\Lambda$ CDM.

## CONTENTS

I. Introduction	4	F. Results: binned $\sigma_8(z)$	25
II. Data and baseline modeling	5	G. Impact of model and analysis choices on $S_8$	26
A. Source and lens catalogues	5	H. Impact of model choice on $\sum m_\nu$	27
B. 2PCF – measurements	5	I. Tension between DES and external data for extended models	28
C. 2PCF – baseline modeling	6	J. Assessing the preference for extended models relative to $\Lambda$ CDM	30
D. Scale cuts	8	VII. Conclusions	31
E. Parameter space	8	Acknowledgements	32
III. Beyond- $\Lambda$ CDM models	9	References	33
A. Dark energy: $w_0 - w_a$	9	A. HALOFIT validation for $w_0 - w_a$	37
B. Curvature: $\Omega_k$	10	B. Validation of Weyl potential pipeline	37
C. Extra relativistic degrees of freedom: $N_{\text{eff}}$	10	C. The FASTISMORE framework for robustness tests	38
D. Massive sterile neutrinos: $N_{\text{eff}} - m_{\text{eff}}$	10	D. Additional validation plots and discussion	39
E. Test of gravity on cosmological scales: $\Sigma_0 - \mu_0$	11	1. All-data robustness plots	39
F. Binned $\sigma_8(z)$	12	2. Robustness investigation: $N_{\text{eff}} - m_{\text{eff}}$	39
IV. External data	13	3. Robustness investigation: binned $\sigma_8(z)$	41
A. Cosmic microwave background	14	4. Hyperrank discussion	41
B. Baryon acoustic oscillations and redshift space distortions	14	E. Metrics for assessing tension between datasets	42
C. Supernovae	15	F. Model comparison metrics	43
V. Analysis procedure and validation	15	1. Bayesian quantities	43
A. Fiducial synthetic data	15	2. MAP-based statistics	44
B. ‘Contaminated’ synthetic data	16	G. Testing beyond- $\Lambda$ CDM model response to $X_{\text{Lens}}$	45
C. Robustness to model variations	17	H. Comparison of Y1 and Y3 $w_0 - w_a$ constraints	45
VI. Results	20		
A. Results: $w_0 - w_a$	20		
B. Results: $\Omega_k$	21		
C. Results: $N_{\text{eff}}$	21		
D. Results: $N_{\text{eff}} - m_{\text{eff}}$	23		
E. Results: $\Sigma_0 - \mu_0$	23		

## I. INTRODUCTION

The discovery of the accelerated expansion of the universe made about two decades ago [1, 2] established  $\Lambda$ CDM as the standard model in cosmology. This paradigm relies on three pillars: that general relativity correctly describes gravitational interactions at cosmological scales; that at those scales the Universe appears homogeneous, isotropic and spatially flat; and that the Universe’s content at late times is dominated by non-relativistic, pressureless cold dark matter (CDM), and the cosmological constant term  $\Lambda$ . The resulting  $\Lambda$ CDM model is in good agreement with cosmological observations from a wide range of temporal and spatial scales [3–22].

The impressive phenomenological success of the  $\Lambda$ CDM model has not been matched in our understanding of the physical nature of dark energy [23, 24], nor in insights as to why the cosmological constant appears to be so small relative to natural scales in particle physics [25–28]. Therefore, cosmology is in need of new and better data that can help shed light on these cosmological conundrums. The quest to understand dark energy has spawned a worldwide effort to measure the growth and evolution of cosmic structures in the Universe. Ongoing experiments focused on dark energy include wide field photometric surveys such as the Dark Energy Survey (DES)<sup>1</sup> [29–31], the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP)<sup>2</sup> [18, 32], the Kilo-Degree Survey (KiDS)<sup>3</sup> [19, 33], in addition to ongoing spectroscopic surveys like the Extended Baryon Oscillation Spectroscopic Survey (eBOSS)<sup>4</sup> [34] and the Dark Energy Spectroscopic Instrument (DESI)<sup>5</sup> [35]. These surveys have demonstrated the feasibility of ambitious large-scale structure analyses, featured development of state-of-the-art systematics calibration, and established new standards in protecting analyses against observer bias before the results are revealed. Thus far, these surveys have provided constraints consistent with the  $\Lambda$ CDM model, and contributed to tightening the constraints on several key cosmological parameters.

Using data from these surveys to search for deviations from the predictions of  $\Lambda$ CDM is one of the primary goals of modern cosmology. Such deviations could provide clues as to where that minimal cosmological model needs to be extended, and thus a deeper understanding of the fundamental physics impacting the large-scale properties of the Universe. One approach to testing the  $\Lambda$ CDM model is to compare  $\Lambda$ CDM parameter estimates inferred from different sets of observables. This is the motivation behind the ongoing exploration of the 3–5 $\sigma$  tension in measurements of the Hubble constant,  $H_0$ , between low-redshift distance-ladder measurements and those from the CMB at  $z \approx 1100$  (see Refs. [36–41] for a summary), as well as the scrutiny of 1–3 $\sigma$  offsets between large scale structure [16–19, 30, 42–46] and CMB-

based [11, 12] constraints on the amplitude of matter density fluctuations scaled by the square root of the matter density,  $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ .

In the present analysis we adopt a complementary approach by constraining cosmological models which add physics beyond that of the standard  $\Lambda$ CDM paradigm. While future precision measurements and careful characterization of existing data (e.g. as in Refs. [47, 48]) will undoubtedly be required to resolve the origin of any tensions between datasets, it is also valuable to investigate whether (or to what extent) any observed offsets may be alleviated by new physics. Additionally, constraining the parameters of extended models can offer greater sensitivity to signatures of beyond- $\Lambda$ CDM physics that may not manifest clearly as a tension between different measurements of  $\Lambda$ CDM parameters.

This paper constrains a range of extended models using a combined analysis of weak-lensing and galaxy-clustering observations from the first three years of data<sup>6</sup> of the Dark Energy Survey (henceforth DES Y3) [49]. The models, are as follows:

- Dynamical dark energy parameterized via the linear expansion of the dark energy equation of state  $w(a) = w_0 + (1 - a)w_a$ ;
- Non-zero spatial curvature  $\Omega_k$ ;
- Varying the effective number of relativistic species  $N_{\text{eff}}$ ;
- Sterile neutrinos varying parameters  $N_{\text{eff}}$  and  $m_{\text{eff}}$  to control the particles’ temperature and effective mass respectively;
- Deviations from General Relativity introduced via the functions  $\Sigma(k, z) = \Sigma_0 \Omega_\Lambda(z)/\Omega_{\Lambda,0}$  and  $\mu(k, z) = \mu_0 \Omega_\Lambda(z)/\Omega_{\Lambda,0}$  respectively modifying the lensing and Poisson equations;
- Variation of the growth rate of structure parameterized by independent  $\sigma_8$  values in different redshift bins.

The selection of these models dictated both by their interest to the cosmology community (this is the primary motivation for  $w_0 - w_a$ ,  $\Omega_k$ ,  $N_{\text{eff}}$ ), and by the kinds of physics about which DES measurements add qualitatively new information (primarily motivating  $N_{\text{eff}} - m_{\text{eff}}$ ,  $\Sigma_0 - \mu_0$ , binned  $\sigma_8(z)$ ).

This work is a successor to the DES Y1 extended-model analysis [50], which for conciseness we will reference as DES-Y1Ext, and complements the main DES Y3 galaxy clustering and weak lensing analysis [44] (henceforth DES-Y3KP) that presented cosmology results for  $\Lambda$ CDM and the  $w$ CDM model testing for a constant dark energy equation of state different from  $-1$ . All of these studies extract cosmological information from DES data via a so-called ‘3 $\times$ 2pt’

<sup>1</sup> <http://www.darkenergysurvey.org/>

<sup>2</sup> <https://www.naoj.org/Projects/HSC/>

<sup>3</sup> <http://kids.strw.leidenuniv.nl/>

<sup>4</sup> <https://www.sdss.org/surveys/eboss/>

<sup>5</sup> <https://www.desi.lbl.gov/>

<sup>6</sup> Publicly available at: <https://des.ncsa.illinois.edu/releases/y3a2>

analysis, in which parameter estimation is based on the combined analysis of three types of projected two-point correlation functions: cosmic shear measurements capturing weak lensing distortions to the shape of background source galaxies, galaxy clustering measurements of the positions of foreground lens galaxies, and the tangential shear of source galaxy shapes around each of the lens positions. Compared to DES-Y1Ext, this work includes a number of updates, the most notable being that the DES Y3 data cover roughly three times the sky area included in the Y1 analysis. To maximize the constraining power of our cosmological data, we will additionally combine the DES Y3  $3 \times 2$ pt constraints with the following external datasets: baryon acoustic oscillations (BAO) and redshift-space distortion (RSD) measurements from the eBOSS, 6dF, and MGS galaxy surveys [15], the Pantheon type Ia supernova (SN) catalogue [8], and the Planck 2018 cosmic microwave background (CMB) data [51].

The paper is organized as follows: In Sec. II we describe the DES Y3 data used in this analysis, and the baseline modeling of the observables. Sections III and IV are devoted to a presentation of the extended models and the main datasets exploited in this work, respectively. In Sec. V we discuss the details of our analysis validation. We present our main results in Sec. VI, and conclude in Sec. VII.

Data supplementing this paper, including chains, scale cuts, and numerical versions of summary plots, will be available online as part of the DES Y3 data release.<sup>7</sup>

## II. DATA AND BASELINE MODELING

In this section we describe the DES data used in this analysis and the likelihood used to perform parameter estimation based on the angular two-point correlation function (2PCF) summary statistics into which those data are condensed.

### A. Source and lens catalogues

The Dark Energy Survey (DES) is a  $5000 \text{ deg}^2$  photometric galaxy survey which, over the course of six years, collected data using the Dark Energy Camera (DECam [52]), mounted on the Víctor Blanco 4m telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. In this work we employ data from the first three years of DES observations (DES Y3), which constitute the DES Data Release 1 (DR1 [49]). That data was processed to produce a photometric catalog of 399 million objects with signal-to-noise ratio of  $\sim 10$  in  $r, i, z$  co-add images. For cosmological inference, we further refine this catalog to produce a ‘Gold’ sample [53] containing 319 million objects, extending to a limiting magnitude of 23 in the  $i$ -band.

From the Gold sample galaxies we select two samples: ‘source’ (background) galaxies, whose shear is used for measurements of gravitational lensing, and ‘lens’ (foreground) galaxies, whose positions are recorded and used for measurements of galaxy clustering. The source galaxy sample is used to produce the DES Y3 shape catalogue [54]. We measure galaxy shapes with the METACALIBRATION pipeline [55, 56], which uses  $r, i, z$ -band information to infer objects’ ellipticity and other photometric properties, employing updates to PSF solutions [57], astrometric solutions [53], and inverse-variance weighting for the galaxies to improve upon a similar pipeline used for the DES Y1 analysis [30]. Multiplicative shear bias is calibrated using image simulations [58] and redshifts are inferred using a self-organizing-map approach [59, 60] which connects wide and deep-field [61] galaxy measurements using Balrog simulations [62]. The final DES Y3 shape catalog contains 100 million galaxies covering an area of  $4143 \text{ deg}^2$ , with a weighted effective number density  $n_{\text{eff}} = 5.9 \text{ per arcmin}^2$  and corresponding shape noise  $\sigma_e = 0.26$ .

The lens galaxy sample that we use in this paper, MagLim [63], is one of the two lens samples considered in DES-Y3KP. The MagLim sample contains sources selected to be in the redshift range  $0.2 \leq z \leq 1.05$ , and is divided into six tomographic bins with nominal edges  $z = [0.20, 0.40, 0.55, 0.70, 0.85, 0.95, 1.05]$ . Uncertainties in the photometric redshift estimator used to define these bins cause the bins’ actual  $n(z)$  redshift distributions to extend outside those bounds. The inferred lens redshift distributions have been further validated using cross-correlations between galaxies in MagLim with spectroscopic galaxy samples [64]. Weights based on the correlation between number density with survey properties mitigate the impact of observing systematics [65]. Refs. [65, 66] describe the sample’s validation and characterization in more detail.

We follow DES-Y3KP in removing the two highest redshift MagLim bins from our analysis, as studies after unblinding the  $\Lambda/w$ CDM results revealed issues with the sample at  $z > 0.85$ . These issues, which manifest as an inability of the model to consistently match the clustering amplitude of galaxy clustering and galaxy–galaxy lensing in the last two tomographic bins, were initially detected because they contribute to poor  $\Lambda/w$ CDM model fits. While this might be of particular interest to searches for beyond- $\Lambda$ CDM physics, investigation since suggests that the issue is most plausibly caused by systematics related to photometric calibration. Further discussion can be found in Refs. [44, 46, 66]. Given these indications, we choose to adopt a conservative approach of removing the impacted MagLim bins from our analysis.

### B. 2PCF – measurements

The cosmological information contained in the lens and source samples described above is then summarized in three two-point correlation functions (2PCF):

- Galaxy clustering: the auto-correlation of lens galaxy positions  $w^i(\theta)$  in each redshift bin  $i$ , i.e. the fractional

<sup>7</sup> <https://des.ncsa.illinois.edu/releases/y3a2/Y3key-extensions>

excess number of galaxy pairs of separation  $\theta$  relative to the number of pairs of randomly distributed points within our survey mask [65],

- **Cosmic shear:** the auto-correlation of source galaxy shapes within and between source redshift bins  $i, j$ , of which there are two components  $\xi_{\pm}^{i,j}(\theta)$ , taking the products of the ellipticity components of pairs of galaxies, either adding (+) or subtracting (−) the component tangential to the line connecting the galaxies and the component rotated by  $\pi/4$  [67, 68],
- **Galaxy–galaxy lensing:** the mean tangential ellipticity of source galaxy shapes around lens galaxy positions  $\gamma_t^{i,j}(\theta)$ , for each pair of redshift bins  $i, j$  [69].

Details of these measurements and the checks for potential systematic effects in them are described in detail in Refs. [65, 67–69], and an overview of the full data vector is given in Ref. [44]. We follow DES-Y3KP and refer to the combined list of  $\{w^i(\theta), \xi_{\pm}^{ij}(\theta), \gamma_t^{ij}(\theta)\}$ , for all angles  $\theta$  and redshift bins  $i$  and  $j$ , as the ‘data vector’. Section II C below has more details about the component pieces of the data vector.

Each of these measurements is performed in a set of 20 logarithmic bins of angular separation between  $2.5'$  and  $250'$  using the software TREECORR [70]. We only use a subset of these bins, removing angular scales where our model is not sufficiently accurate, as discussed in Sec. II D.

### C. 2PCF – baseline modeling

Our baseline modeling methodology generally follows that used in DES-Y3KP [44], and described in detail in the methodology Y3 paper [71]. Notable differences from the DES-Y3KP analysis include the use of a simpler non-linear alignment (NLA) intrinsic alignment model as opposed to the tidal alignment and tidal torquing (TATT) model as our fiducial intrinsic alignment model, and not using the shear-ratio likelihood [72] in most of our analysis. Below we summarize the modeling used to compute the  $3 \times 2$ pt likelihood, as well as these differences.

As noted above, the Y3- $3 \times 2$ pt analysis consists of a set of 2PCF measurements describing the angular correlation of lens galaxy positions and source galaxy shapes for several redshift bins. We model the likelihood as Gaussian in the data vector  $\mathbf{D}$ ,

$$\ln \mathcal{L}(\mathbf{D}|\Theta) = -\frac{1}{2} \left[ (\mathbf{D} - \mathbf{M}(\Theta))^T \mathbf{C}^{-1} (\mathbf{D} - \mathbf{M}(\Theta)) \right] + L_0, \quad (1)$$

where  $\Theta$  is the vector of cosmological and nuisance parameters,  $\mathbf{C}$  is the covariance, and  $L_0$  is a normalization constant. The covariance is computed analytically using COSMOLIKE [71] including COSMOCOV [73]. The likelihood and the covariance were validated in Ref. [74], where it has been shown that, for the precision level attained by the DES Y3 analysis, assuming a Gaussian likelihood with the various assumptions involved in the computation of  $\mathbf{C}$  are all excellent approximations (see in particular Fig. 1 of Ref. [74]). We sample the

above likelihood to obtain posterior and evidence estimates using the POLYCHORD nested sampler [75, 76], following guidelines for settings described in Ref. [77]. The length of the fiducial data vector  $\mathbf{D}$  is 462 though for some models data points will be removed to account for modeling uncertainties (see Table I). The length of the parameter vector  $\Theta$  for  $\Lambda$ CDM is 28 when fitting DES data alone (additional parameters are introduced when combining with external data).

Full details of how the data vector  $\mathbf{D}$  is theoretically predicted can again be found in Refs. [44, 71], but here we give a brief overview. The 2PCF are computed from the observed projected galaxy density contrast  $\delta_g$  and the shear field decomposed into E- and B-modes. The 2PCF forming the data vector  $\mathbf{D}$  can be expressed in terms of the angular power spectra as:

$$\begin{aligned} w^i(\theta) &= \sum_{\ell} \frac{2\ell+1}{4\pi} P_{\ell}(\cos \theta) C_{\delta_g \delta_g}^{ii}(\ell) \\ \gamma_t^{ij}(\theta) &= \sum_{\ell} \frac{2\ell+1}{4\pi} \frac{P_{\ell}^2(\cos \theta)}{\ell(\ell+1)} C_{\delta_g E}^{ij}(\ell) \\ \xi_{\pm}^{ij}(\theta) &= \sum_{\ell \geq 2} \frac{2\ell+1}{4\pi} \frac{2(G_{\ell,2}^{+}(x) \pm G_{\ell,2}^{-}(x))}{\ell^2(\ell+1)^2} \\ &\quad \times [C_{EE}^{ij}(\ell) \pm C_{BB}^{ij}(\ell)], \end{aligned} \quad (2)$$

where  $i, j$  denote redshift bins. Here  $P_{\ell}$  are the Legendre polynomials of order  $\ell$ ,  $P_{\ell}^m$  are the associated Legendre polynomials,  $x = \cos \theta$  and the functions  $G_{\ell,m}^{+/-}(x)$  are a combination of the associated Legendre polynomials  $P_{\ell}^m(x)$  and  $P_{\ell-1}^m(x)$  and are given explicitly in Eq. (4.19) of Ref. [78]. Following DES-Y3KP, we only consider the auto-correlations  $w^i(\theta)$  for each tomographic bin since in the DES Y1 and Y3 analyses it was shown that the cross correlations do not add much constraining power and would make our analysis much more susceptible to systematic errors related to the modeling of magnification and redshift distributions [14, 63].

The angular power spectra  $C(\ell)$  that enter Eqs. (2) combine integrals over tracer distributions with astrophysical contributions from intrinsic alignments (IA), magnification, and redshift space distortions (RSD). The shear-shear (EE, BB) and galaxy-shear ( $\delta_g E$ ) spectra are computed using the Limber approximation and include magnification and IA, but neglect RSD. Galaxy-galaxy clustering ( $\delta_g \delta_g$ ) is computed with via non-Limber integrals with contributions from both RSD and magnification. These calculations are described below.

Using the Limber approximation [79] and assuming spatial flatness, the angular power spectra for for cosmic shear and galaxy-galaxy lensing can be written in a general form

$$C_{AB}^{ij}(\ell) = \int d\chi \frac{q_A^i(\chi) q_B^j(\chi)}{\chi^2} P_{\delta} \left( k = \frac{\ell + 0.5}{\chi}, z \right), \quad (3)$$

where  $\{A, B\} \in \{\delta_g, \kappa\}$  and  $\kappa$  is the weak lensing convergence whose contributions to shear correlations will be detailed below. Here,  $P_{\delta}(k, z)$  is the three-dimensional matter power spectrum evaluated at wavenumber  $k$  and redshift  $z$ . We use CAMB to compute the linear  $P_{\delta}(k, z)$  and

HALOFIT [80–82] to do nonlinear modeling. The radial weight functions  $q_{A,B}^i$  are given by

$$q_{\kappa}^i(\chi) = \frac{3H_0^2\Omega_m\chi}{2a(\chi)} \int_{\chi}^{\chi_h} d\chi' \left( \frac{\chi' - \chi}{\chi'} \right) n_{\kappa}^i(z(\chi')) \frac{dz}{d\chi'} \\ q_{\delta_g}^i(\chi) = b^i n_{\delta}^i(z(\chi)) \frac{dz}{d\chi}. \quad (4)$$

Here  $H_0$  is the Hubble parameter today,  $\Omega_m$  is the ratio of today's matter density to today's critical density, and  $a(\chi)$  is the Universe's scale factor at comoving distance  $\chi$ . For conciseness, we refer Refs. [83, 84] for the full non-Limber expressions used for  $C_{\delta_g\delta_g}^{ij}(\ell)$ .

In the expression above, we adopt a linear galaxy bias model to relate the galaxy density  $\delta_g$  to the matter density:  $\delta_g = b^i \delta$ , with  $b^i$  the galaxy bias in lens redshift bin  $i$ , which we vary in the analysis. Furthermore,  $n_{\kappa}^i(z)$  and  $n_{\delta}^i(z)$  denote the redshift distributions of the different DES Y3 redshift bins of source and lens galaxies respectively, normalized so that  $\int dz n_{\kappa,\delta}^i(z) = 1$ .

Contributions to observed spectra from intrinsic alignments and magnification are included as follows:

$$C_{EE}^{ij}(\ell) = C_{\kappa\kappa}^{ij}(\ell) + C_{\kappa I_E}^{ij}(\ell) + C_{\kappa I_E}^{ji}(\ell) + C_{I_E I_E}^{ij}(\ell) \\ C_{BB}^{ij}(\ell) = C_{I_B I_B}^{ij}(\ell), \\ C_{\delta_g E}^{ij}(\ell) = C_{\delta_g \kappa}^{ij}(\ell) + C_{\delta_g I_E}^{ij}(\ell) + C_i C_j C_{\kappa\kappa}^{ij}(\ell) + C_i C_{\kappa I_E}^{ij}(\ell) \\ C_{\delta_g \delta_g}^{ii}(\ell) = C_{\delta_g \delta_g}^{ii}(\ell) + 2C_i C_{\delta_g \kappa}^{ii}(\ell) + C_i^2 C_{\kappa\kappa}^{ii}(\ell). \quad (5)$$

Here  $I_{E/B}$  refers to the E/B-modes of the intrinsic alignment (IA), the prime denotes non-magnified power spectra, and  $C_i$  are magnification constants. We emphasize that compared to the calculations that we employ in practice, the expression for galaxy clustering in Eq. (5) neglects contributions from RSD. These RSD contributions, which depend on the linear growth factor  $f(\chi)$ , are incorporated — along with magnification — in the non-Limber calculation of  $C_{\delta_g\delta_g}^{ii}(\ell)$ .

We now describe the IA and magnification effects, as well as their modeling, in more detail. IA refers to the fact that galaxies tend to align because of their gravitational environment, thus contributing to the cosmic shear signal. Here, we adopt the non-linear tidal alignment model as our fiducial IA model [85, 86]. The NLA model assumes that intrinsic galaxy shapes are linearly proportional to the fully nonlinear tidal field, calculated using the nonlinear power spectrum. While this ansatz is not a fully consistent nonlinear model, it is straightforward to calculate and has been shown to more accurately describe observed IA than linear theory (see, e.g., Ref. [87]). Our fiducial NLA model has two free parameters,  $a$  and  $\eta$ , which control the amplitude and redshift dependence of IA, respectively. IA includes both gravitational lensing–intrinsic (GI) and intrinsic–intrinsic (II) contributions, whose power spectra are then given by

$$P_{GI}(k, z) = A_1(z) P_{\delta}(k, z), \quad P_{II}(k, z) = A_1^2(z) P_{\delta}(k, z). \quad (6)$$

The pre-factor  $A_1(z)$  is

$$A_1(z) = -a \bar{C}_1 \frac{\rho_{\text{crit}} \Omega_m}{D(z)} \left( \frac{1+z}{1+z_0} \right)^{\eta}, \quad (7)$$

where  $D(z)$  is the linear growth factor normalized to be equal to  $(1+z)^{-1}$  at high redshifts,  $\rho_{\text{crit}}$  is the critical density and  $\bar{C}_1$  is a normalisation constant, by convention fixed at  $\bar{C}_1 = 5 \times 10^{-14} M_{\odot}^{-1} h^{-2} \text{Mpc}^3$ . The IA angular power spectra are then computed using Eq. (3) with the kernel  $q_{\kappa}^i(\chi) = n_{\kappa}^i(z(\chi)) dz/d\chi$  (see Ref. [68] for more detailed discussion of the IA modeling and implementation for DES Y3 cosmic shear).

Our decision to adopt the NLA model contrasts with the DES-Y3KP  $\Lambda/w$ CDM analysis, which adopted a more complicated tidal alignment and tidal torquing (TATT) IA model. The systematic tests carried out prior to the DES-Y3KP analysis motivated the use of the TATT model because when analyzing synthetic data containing tidal torquing effects of a size allowed by DES Y1 constraints, cosmological constraints using the simpler NLA model were found to be biased. However, the  $\Lambda$ CDM analysis of the Y3-3 $\times$ 2pt data has subsequently shown preference for a generally lower amplitude of intrinsic alignments, finding that the NLA model is sufficient for unbiased modeling at the Y3 precision level. With the benefit of these  $\Lambda$ CDM results, and the desire to limit the number of nuisance parameters in our extended-model analysis, we thus opt for the NLA model. We do however run additional chains that use TATT, and prior to unblinding we check if there is a preference for the TATT model over NLA in any of the beyond- $\Lambda$ CDM models. We use the Fast-PT code [88] implemented in COSMOSIS in order to compute both the NLA and TATT contributions, unless specified otherwise.

As noted above, we include the contribution to galaxy clustering from the magnification of the lens sample density in Eq. (5) using magnification constants  $C_i$ . These constants are determined by the selection function of the lens sample tomographic bin such that the magnified number density is related to the convergence experienced by lens galaxies through:  $\delta_{g,\text{mag}}^i = C_i \kappa^i$ . The prime in Eq. (5) indicates the power spectrum unmodified by magnification. We fix the coefficients  $C_i$  to values indicated in Table II, which were determined in Ref. [89] using the Balrog image simulations [62]. Note that in V we test the sensitivity of our cosmology results to inaccuracies in these assumed values.

The baseline analysis of DES-Y3KP includes a shear ratio likelihood. This quantity incorporates the ratio between  $\gamma_t$  measurements with the same lens bin and different shear bins [72]. While previously studied in the context of constraining dark energy models [90], it has more recently been found that shear ratio's particular strength is its sensitivity to the redshift distribution of source galaxies [91]. In all model extensions other than binned  $\sigma_8(z)$ , we do not include this shear ratio likelihood. Recall that the motivation for including shear ratio is to add additional geometric constraining power which for instance helps reduce photometric redshift uncertainties. However, simulated analyses for our extended model showed that the inclusion or not of the shear ratio likelihood had a minimal impact on 3 $\times$ 2pt constraints. Given this, and the

Scale cuts	Data points					Used for extended models
	$\xi_+$	$\xi_-$	$\gamma_t$	$w$	Total	
Fiducial	166	61	192	43	462	$w_0 - w_a, N_{\text{eff}}, \text{binned } \sigma_8(z)$
Linear	105	3	105	43	256	$N_{\text{eff}} - m_{\text{eff}}, \Sigma_0 - \mu_0$
Linear+Limber	100	2	100	19	221	$\Omega_k$

TABLE I. Number of  $3 \times 2$ pt data points remaining after the different sets of scale cuts used in this analysis. The fiducial cuts are the same as those used in DES-Y3KP, linear cuts remove additional points at small scales affected by nonlinear structure growth, and linear+Limber cuts remove data points both at nonlinear scales and where non-Limber calculations are needed to accurately model large-angle galaxy clustering. Unless otherwise noted, whenever a comparison is shown between an extended model and  $\Lambda$ CDM, the  $\Lambda$ CDM results will use scale cuts matching those of the extended model.

lack of extended modeling validation of that likelihood, we have opted to not include it as part of our baseline analysis.

#### D. Scale cuts

As in the DES-Y3KP analysis, we define scales below which we remove measurements from our analysis to mitigate the limits of the  $3 \times 2$ pt modeling. Modeling uncertainties of measurements at small angular scales may otherwise lead to systematic biases in cosmological parameter estimates. We refer to this approach as scale cuts. In the end the likelihood calculation in Eq. (1) only uses 2PCF measurements that remain after such cuts. Our baseline scale cuts are the same as those used for the  $\Lambda$ CDM analysis of DES-Y3KP. As described in detail in the DES Y3 methods paper [71], these cuts were defined based on the iterative analysis of synthetic data. Specifically, that data was a theoretical prediction of the 2PCF observables that included two significant systematic effects not included in our model: baryonic feedback effects extracted from the OWLS AGN hydrodynamic simulations and non-linear galaxy bias. By repeatedly analyzing that synthetic data while removing successively more small-angle data points, we determined scale cuts at which the biases on  $\Omega_m$  and  $S_8$  due to each of the unmodeled systematics were below  $0.3\sigma$ . This determines the fiducial scales used for the DES Y3  $3 \times 2$ pt analysis, where the number of data points for each of the 2PCF is summarized in the first line of Table I.<sup>8</sup> These same cuts are used for the  $w$ CDM,  $w_0 - w_a$ ,  $N_{\text{eff}}$  and binned  $\sigma_8(z)$  models.

For several of the models studied in this paper, we have chosen a stricter set of scale cuts than the fiducial case. Specifically, for models with nonzero  $\Omega_k$ , at the time of this analysis HALOFIT had not been sufficiently validated on non-linear scales<sup>9</sup>; for  $N_{\text{eff}} - m_{\text{eff}}$  models, HALOFIT is known to be

not sufficiently well calibrated; and finally the  $\Sigma_0 - \mu_0$  tests of gravity are only well-defined on linear scales. For these classes of models, we restrict our analysis to purely linear scales (for nonzero curvature there will be an additional scale cut, discussed below). To determine those scales, we follow the procedure first applied in the *Planck* 2015 analysis [93] and followed later in DES-Y1Ext. We compute the difference between the nonlinear and linear theory predictions of the 2PCF in the standard  $\Lambda$ CDM model at a fiducial cosmology on scales left after fiducial scale cuts. Using the respective data vector theory predictions,  $\mathbf{D}_{\text{NL}}$  and  $\mathbf{D}_{\text{lin}}$ , and full error covariance of DES Y3,  $\mathbf{C}$ , we calculate the quantity

$$\Delta\chi^2 \equiv (\mathbf{D}_{\text{NL}} - \mathbf{D}_{\text{lin}})^T \mathbf{C}^{-1} (\mathbf{D}_{\text{NL}} - \mathbf{D}_{\text{lin}}) \quad (8)$$

and identify the single data point that contributes most to this quantity. We remove that data point, and repeat the process until  $\Delta\chi^2 < 1$ . This constitutes our set of linear scales, used for  $N_{\text{eff}} - m_{\text{eff}}$  and  $\Sigma_0 - \mu_0$ . The resulting linear scale cuts lead to a  $3 \times 2$ pt data vector of 256 elements (see second line of Table I).

The third and final choice of scale cuts removes both scales which are impacted by nonlinear structure growth, and those requiring non-Limber projection calculations to accurately model galaxy clustering. This is relevant for the  $\Omega_k$  analysis, because the accuracy of the fast non-Limber method [83] used to model large-angle galaxy clustering has not been tested for  $\Omega_k \neq 0$ . The procedure to identify these scales is the same as that described for linear scale cuts above, except that the synthetic data vector  $\mathbf{D}_{\text{lin}}$  is replaced with  $\mathbf{D}_{\text{lin+Limber}}$ , which, in addition to having only linear modeling for the matter power spectrum, is computed using the Limber approximation at all angular scales. The linear+Limber cuts are thus slightly more stringent than the linear-only cuts. Table I shows that the resulting cuts lead to a  $3 \times 2$ pt data vector of 221 elements that are then used in the  $\Omega_k$  analysis. Note that the iterative nature of our scale-cut definition causes the Linear+Limber cuts to remove additional points from  $\xi_{\pm}(\theta)$  and  $\gamma_t(\theta)$  compared to the Linear cuts, even though non-Limber calculations are only used for galaxy clustering  $w(\theta)$  calculations.

#### E. Parameter space

The parameters and their priors used in our baseline analysis match those of DES-Y3KP. The cosmological parameters are

$$\Theta_{\text{base}} = \{\Omega_m, \Omega_b, A_s, n_s, h, \Omega_\nu h^2\} \quad (9)$$

(equivalently,  $\Omega_\nu h^2$  can be replaced by  $\sum m_\nu$ ), with the neutrinos modeled as three degenerate species with equal masses. The priors on these parameters are listed in the top section of Table II. We adopt an additional prior requiring  $0.005 < \Omega_b h^2 < 0.040$ . This baryon-density prior is introduced because we include a Big Bang Nucleosynthesis (BBN) consistency condition which imposes a relation between the physical baryon density  $\Omega_b h^2$ , the relativistic degrees of freedom

<sup>8</sup> Minimum and maximum scales used after the scale cuts procedure are indicated in the COSMOSIS files shared as part of the data release.

<sup>9</sup> Ref. [92], which was released when this paper was in final stages of preparation, represents a promising approach to improve this.

Parameter	Prior	
<b>Base Cosmology</b>		
$\Omega_{\text{m}}$	Flat	(0.1, 0.9)
$10^9 A_{\text{s}}$	Flat	(0.5, 5.0)
$n_{\text{s}}$	Flat	(0.87, 1.07)
$\Omega_{\text{b}}$	Flat	(0.03, 0.07)
$h$	Flat	(0.55, 0.91)
$10^3 \Omega_{\nu} h^2$	Flat	(0.60, 6.44)
		$0.005 < \Omega_{\text{b}} h^2 < 0.040$
<b>Extended Cosmology</b>		
$w_0, w_a$	Flat	$w_0 \in (-3.0, -0.33)$ $w_a \in (-3.0, 3.0)$ $w_0 + w_a < 0$
$\Omega_k$	Flat	(-0.25, 0.25)
$N_{\text{eff}}$	Flat	(1.0, 10.0)
$N_{\text{eff}}, m_{\text{eff}}$	Flat	$N_{\text{eff}} \in (3.044, 10)$ $m_{\text{eff}} \in (0.0, 3.0) \text{ eV}$
$\Sigma_0, \mu_0$	Flat	$\Sigma_0, \mu_0 \in (-1.5, 1.5)$ $\mu_0 > 2\Sigma_0 + 1$
$A_i^{P_{\text{lin}}}(i \in [2, 4])$	Flat	(0.1, 3)
<b>Lens Galaxy Bias</b>		
$b_i(i \in [1, 4])$	Flat	(0.8, 3.0)
<b>Lens magnification</b>		
$\mathcal{C}_1$	Fixed	0.42
$\mathcal{C}_2$	Fixed	0.30
$\mathcal{C}_3$	Fixed	1.76
$\mathcal{C}_4$	Fixed	1.94
<b>Lens photo-z</b>		
$\Delta z_1^1 \times 10^2$	Gaussian	(-0.9, 0.7)
$\Delta z_1^2 \times 10^2$	Gaussian	(-3.5, 1.1)
$\Delta z_1^3 \times 10^2$	Gaussian	(-0.5, 0.6)
$\Delta z_1^4 \times 10^2$	Gaussian	(-0.7, 0.6)
$\sigma_{z,1}^1$	Gaussian	(0.98, 0.06)
$\sigma_{z,1}^2$	Gaussian	(1.31, 0.09)
$\sigma_{z,1}^3$	Gaussian	(0.87, 0.05)
$\sigma_{z,1}^4$	Gaussian	(0.92, 0.05)
<b>Intrinsic Alignment</b>		
$a$	Flat	(-5, 5)
$\eta$	Flat	(-5, 5)
<b>Source photo-z</b>		
$\Delta z_{\text{s}}^1 \times 10^2$	Gaussian	(0.0, 1.8)
$\Delta z_{\text{s}}^2 \times 10^2$	Gaussian	(0.0, 1.5)
$\Delta z_{\text{s}}^3 \times 10^2$	Gaussian	(0.0, 1.1)
$\Delta z_{\text{s}}^4 \times 10^2$	Gaussian	(0.0, 1.7)
<b>Shear calibration</b>		
$m^1 \times 10^2$	Gaussian	(-0.6, 0.9)
$m^2 \times 10^2$	Gaussian	(-2.0, 0.8)
$m^3 \times 10^2$	Gaussian	(-2.4, 0.8)
$m^4 \times 10^2$	Gaussian	(-3.7, 0.8)
<b>External data</b>		
$\tau$ ( <i>Planck</i> )	Flat	(0.01, 0.8)
$A_{\text{P}}$ ( <i>Planck</i> )	Gaussian	(1.0, 0.0025)
$M$ (SN)	Flat	(-20, -18)

TABLE II. Parameters and priors describing the baseline cosmology, extended models and nuisance parameters used in this analysis. We quote the lower and upper limits of flat priors and the mean and standard deviation of Gaussian priors. The parameter  $w$  is fixed to  $-1$  for all models other than  $w\text{CDM}$  and  $w_0 - w_a$ , and for  $w\text{CDM}$  it uses the same prior as for  $w_0$ . For the  $N_{\text{eff}} - m_{\text{eff}}$  model we fix the sum of active neutrino masses to 0.06 eV.

$N_{\text{eff}}$ , and the Helium abundance  $Y_{\text{He}}$  [94].<sup>10</sup> This consistency relation only alters calculations when  $N_{\text{eff}}$  is varied, but introduces the  $\Omega_b h^2$  prior for all models because it relies on a table defined for a finite range of physical baryon density and so rejects samples outside that range.

We also vary a number of nuisance parameters to describe systematic effects. The intrinsic alignment is described by two parameters,  $a$  and  $\eta$ , and the linear galaxy bias by one parameter  $b_i$  for each of the lens bins; these parameters are assigned flat priors. Additionally, each lens bin has two nuisance parameters: one that controls the mean of the photometric redshift distribution in redshift bin  $i$ ,  $\Delta z_1^i$ , and another which stretches or compresses the  $n(z)$  distribution in  $z$ ,  $\sigma_{z,1}^i$ . Each of the four source bins has a photometric redshift uncertainty parameter  $\Delta z_s^i$ , as well as a shear calibration parameter  $m^i$ .

As indicated at the bottom of Table II, we also vary the optical depth  $\tau$  along with a number of nuisance parameters associated with the *Planck* likelihood, as described in Sec. IV A, when using *Planck* data. When the Pantheon supernovae likelihood is included, we additionally sample over the absolute magnitude of the supernovae  $M$ , as described in Sec. IV C.

### III. BEYOND- $\Lambda\text{CDM}$ MODELS

We now introduce the beyond- $\Lambda\text{CDM}$  models constrained in this paper. For each, we introduce its physics and parameterization, and describe any alterations required to the baseline approach for modeling  $3 \times 2\text{pt}$  observables and performing parameter estimation.

#### A. Dark energy: $w_0 - w_a$

We use the phenomenological model proposed in [95, 96], often referred to as the CPL model, for a time-varying dark energy equation of state:

$$w(a) = w_0 + (1 - a)w_a, \quad (10)$$

where  $a$  is the scale factor, and  $w_0$  and  $w_a$  two new parameters. This is the most commonly considered parameterization of dark energy equation of state with more than one parameter, and has been shown to provide a good fit to a number of dynamical dark energy models that have a more complete physical description [95].

We will also report constraints on the value of  $w(a)$  at the so-called pivot redshift [97]  $z_p = a_p^{-1} - 1$ ,  $w_p \equiv w(a_p)$ . Here  $a_p$  is the scale factor where we have the strongest constraints on  $w(a)$ , and therefore where the value of the equation of state and its derivative with respect to the scale factor are uncorrelated. Using this parameter, Eq. (10) can be rewritten as

$$w(a) = w_p + (a_p - a)w_a. \quad (11)$$

<sup>10</sup> <http://parthenope.na.infn.it/>

The value of the pivot scale factor is determined using the marginalized parameter covariance as

$$a_p = 1 + \frac{C_{w_0 w_a}}{C_{w_a w_a}}. \quad (12)$$

In the flat  $w_0 - w_a$  model, the expansion rate becomes

$$\frac{H^2(a)}{H_0^2} = \Omega_m a^{-3} + (1 - \Omega_m) a^{-3(1+w_0+w_a)} e^{-3w_a(1-a)} \quad (13)$$

As described in Appendix A, we performed additional validation tests to ensure that the use of HALOFIT in our calculation of the nonlinear matter power spectrum is valid for the  $w_0 - w_a$  model. We use the fiducial scale cuts for this model (see Table I).

### B. Curvature: $\Omega_k$

To define the curvature density  $\Omega_k$ , it is most convenient to start from the Friedmann–Lemaître–Robertson–Walker (FLRW) metric in the form

$$ds^2 = -dt^2 + a^2 [d\chi^2 + r^2(\chi) (d\theta^2 + \sin^2 \theta d\phi^2)], \quad (14)$$

where  $\chi$  is the comoving (radial) distance and  $a$  is the scale factor. Then the angular diameter distance  $r(\chi)$  is defined as:

$$r(\chi) = \begin{cases} K^{-1/2} \sin(K^{1/2} \chi) & \text{for } K > 0 \\ \chi & \text{for } K = 0 \\ |K|^{-1/2} \sinh(|K|^{1/2} \chi) & \text{for } K < 0, \end{cases} \quad (15)$$

where  $K$  is the curvature term. A positively curved space has  $K > 0$ , negatively curved corresponds to  $K < 0$ , and flat space has  $K = 0$ . With the curvature of arbitrary sign, the expansion rate can be written as

$$\frac{H^2(a)}{H_0^2} = \Omega_m a^{-3} + (1 - \Omega_m - \Omega_k) + \Omega_k a^{-2}, \quad (16)$$

where  $\Omega_k = -K/H_0^2$ . It then follows that  $\Omega_k < 0$  corresponds to positive spatial curvature, and  $\Omega_k > 0$  to negative.

As noted in Sec. II D, for the curved-universe analysis, due to lack of validated modeling we use a conservative set of scale cuts which avoid both nonlinear scales and the large angular scales where the non-Limber calculation is adopted to model galaxy clustering.<sup>11</sup> For the angular scales where the Limber approximation is used, we apply the commonly-used angular-diameter rescaling approximation for the impact of curvature on line-of-sight projection, which replaces  $\chi$  in Eq. (3) with the angular diameter distance  $r(\chi)$ .

<sup>11</sup> This is a more conservative choice than the approach taken in DES-Y1Ext, where  $\Omega_k$  constraints used the fiducial scale cuts as the Limber approximation was used at all scales.

### C. Extra relativistic degrees of freedom: $N_{\text{eff}}$

We next consider a model that allows for new radiative degrees of freedom in the early Universe, described by the parameter  $N_{\text{eff}}$ . This parameter relates contributions to the energy density in radiation in the early Universe from relativistic species to that of photons via

$$\rho_{\text{rad}} = \left[ 1 + N_{\text{eff}} \frac{7}{8} \left( \frac{T_{\nu,0}}{T_{\gamma,0}} \right)^4 \right] \rho_{\gamma}. \quad (17)$$

Here  $\rho_{\text{rad}}$  and  $\rho_{\gamma}$  are the co-moving energy densities of radiation and photons after electron–positron annihilation. In the standard cosmological model, all contributions to  $N_{\text{eff}}$  come from neutrinos and its value is  $N_{\text{eff}} = 3.044$ , corresponding to three neutrino species plus small corrections due to their non-instantaneous decoupling from photons [98, 99].

We capture the effects of  $N_{\text{eff}}$  by using CAMB’s predictions for its impact on the expansion history and power spectra, using a modified version of the COSMOSIS CAMB interface. We set the CAMB parameters so that each of the three massive neutrino species is assigned a degeneracy [100] of  $\frac{1}{3} N_{\text{eff}}$ . This means that varying  $N_{\text{eff}}$  has a continuous effect on the neutrino temperature  $T_{\nu}$ , with  $\Delta N_{\text{eff}} = 0$  corresponding to the temperature if there are no additional relativistic species beyond the standard model. We do not apply any other modifications to the fiducial model.

### D. Massive sterile neutrinos: $N_{\text{eff}} - m_{\text{eff}}$

We additionally constrain the properties of a light relic particle with non-zero mass, modeled as single species of thermal sterile neutrino. The properties of the sterile neutrino are controlled by the parameters  $N_{\text{eff}}$  and  $m_{\text{eff}}$ . The impact of sterile neutrinos on CMB observables is fairly similar to that of varying  $N_{\text{eff}}$  alone, while its impact on large scale structure has a richer phenomenology. Like active neutrinos, sterile neutrinos suppress large scale structure formation at scales smaller than a free-streaming length scale ( $k > k_{\text{fs}}$ ), with the magnitude of that suppression at high  $k$  controlled by their contribution to cosmological energy density  $\Omega_{\nu_s}$ . The free-streaming scale is set by both the particle’s physical mass and temperature, and the relationship between those properties and the parameters  $N_{\text{eff}}$  and  $m_{\text{eff}}$  depend on the specifics of the model considered. In this analysis, we choose to model the sterile neutrino as a thermal relic, that is, a stable particle species which was once in thermal equilibrium with standard model particles but decoupled at an early time. With this assumption, the particle’s physical mass is  $m_{\text{th}} = m_{\text{eff}} (\Delta N_{\text{eff}})^{-3/4}$  and in linear theory the free-streaming scale is [101]

$$k_{\text{fs}} = \frac{0.8 h \text{Mpc}^{-1}}{\sqrt{1+z}} \left( \frac{m_{\text{eff}}}{(1 \text{eV}) \Delta N_{\text{eff}}} \right). \quad (18)$$

While this thermal model is just one of several possible choices one could make for describing sterile neutrinos, our constraints will represent a more general search for new

physics. As is discussed in Refs. [102?–104], this kind of two-parameter  $N_{\text{eff}} - m_{\text{eff}}$  model is sufficient to perform a generic search for a population of stable, non-interacting massive relic particles. [103, 104].

Here the parameter  $\Delta N_{\text{eff}} \equiv N_{\text{eff}} - 3.044$  determines the temperature of the sterile neutrino, which is related to the standard model temperature of active neutrinos via  $T_{\nu_s} = (\Delta N_{\text{eff}})^{1/4} T_{\nu_a}$ . Thus,  $\Delta N_{\text{eff}} = 1$  corresponds to a sterile neutrino that thermalizes at the same temperature as the active neutrinos, while lower  $N_{\text{eff}}$  means the sterile neutrinos are colder. The parameter  $m_{\text{eff}}$  is an effective mass which captures how the sterile neutrino contributes to the cosmological energy densities, defined so that

$$\Omega_{\nu_s} h^2 = \frac{m_{\text{eff}}}{94.1 \text{ eV}}. \quad (19)$$

When we consider sterile neutrinos the conversion factor between the particle mass and  $\Omega_{\nu} h^2$  is slightly different from the 93.14 eV value used for active neutrinos. This is because sterile neutrinos are assumed not to be affected by electron-positron annihilation in the same way as active neutrinos. Note that both versions of the  $\Omega_{\nu}$ -to-mass conversion factor encode a number of standard model assumptions which cannot be disentangled from cosmological constraints on neutrino mass. Our measurements thus serve as both a test of the mass of neutrinos and of those assumptions. As with the  $N_{\text{eff}}$  model described above, we use CAMB along with a modified version of the COSMOSIS CAMB interface to compute the impact of the sterile neutrinos on expansion history and the linear matter power spectrum.

We assume that active neutrino temperatures are at their standard model value in the instantaneous decoupling approximation,  $T_{\nu_a} = (4/11)^{1/3} T_{\gamma}$ , and following the *Planck* 2018 cosmology analysis [11], we fix the active neutrino mass to the minimum allowed by neutrino oscillation experiments,  $\sum m_{\nu} = 0.06$  eV. Additionally, because the presence of massive light relics like sterile neutrinos complicates the modeling of the nonlinear matter power spectrum as well as galaxy bias [102, 105? , 106] and there are not readily available tools to account for the impact of sterile neutrinos on nonlinear power spectrum modeling (see e.g. Refs. [107, 108]), when constraining  $N_{\text{eff}} - m_{\text{eff}}$  we restrict our analysis to linear scales using the scale-cut procedure described in Sec. II D.

Note that our fiducial prior has a lower bound of  $\Delta N_{\text{eff}} = 0$ , which means our parameter space will include the small- $\Delta N_{\text{eff}}$  regime where the sterile neutrino will be indistinguishable from cold dark matter. As we will find in Sec. V C, including this unconstrained region makes parameter estimation more susceptible to projection effects and thus less robust to the details of nuisance parameter marginalization and the data's noise realization. Given this, in order to obtain a more robust set of constraints and to allow more direct comparison with other studies, we report  $N_{\text{eff}} - m_{\text{eff}}$  constraints using two alternative priors: one where the lower bound of the prior is raised to require  $\Delta N_{\text{eff}} > 0.047$ , corresponding to the minimum temperature for a fermion relic particle that was ever in thermal equilibrium with standard model particles [? ], as well as the same model-specific prior used in *Planck* analyses, requiring  $m_{\text{th}} \leq 10$  eV.

## E. Test of gravity on cosmological scales: $\Sigma_0 - \mu_0$

We test gravity on cosmological scales by adopting the common  $\Sigma, \mu$  phenomenological parameterization proposed and developed in Refs. [109–118]. This model has recently been tested using CMB measurements by the *Planck* satellite and weak lensing data from surveys such as CFHTLenS, KiDS and DES in Refs. [11, 50, 93, 119–121]. In this approach, deviations from the gravitational physics described by General Relativity (GR) are introduced through modifications to the Poisson and lensing equations which then take the following form in Fourier space:

$$\begin{aligned} k^2 \Psi &= -4\pi G a^2 [1 + \mu(a, k)] (\rho\delta + 3(\rho + P)\sigma), \\ k^2 \Phi &= -4\pi G a^2 [1 + \Sigma(a, k)] (2\rho\delta + 3(\rho + P)\sigma). \end{aligned} \quad (20)$$

Here  $\Psi$  is the Newtonian gravitational potential, which determines the gravitational interactions of massive particles,  $\Phi$  is the Weyl potential with which massless particles interact gravitationally,  $\delta$  corresponds to density fluctuations in the comoving gauge,  $\rho$  to matter density, and  $(P + \rho)\sigma$  to the fluid anisotropic stress potential. The functions  $\Sigma(a, k)$  and  $\mu(a, k)$  represent deviations from GR, with  $\Sigma = \mu = 0$  recovering the predictions of GR. This parameterization is equivalent to modifications to the gravitational constant  $G$ , and  $\Sigma/\mu$  are sometimes denoted as  $G_{\Phi/\Psi}$  respectively.

We assume a time dependence following the energy density of the effective dark energy in units of the critical density  $\Omega_{\Lambda}(a)$  normalized by its value today  $\Omega_{\Lambda,0}$ , as done previously in Refs. [11, 50, 109, 119, 120]:

$$\begin{aligned} \Sigma(a, k) &= \Sigma_0 \frac{\Omega_{\Lambda}(a)}{\Omega_{\Lambda,0}}, \\ \mu(a, k) &= \mu_0 \frac{\Omega_{\Lambda}(a)}{\Omega_{\Lambda,0}}. \end{aligned} \quad (21)$$

This parameterization is designed to be sensitive to deviations from GR that are associated with cosmic acceleration. As is pointed out in e.g. Ref. [122], these assumptions may cause our  $\Sigma_0 - \mu_0$  parameterization to lack sensitivity to some modified gravity signals that could be captured by searches with less restrictive assumptions. However, the parameterization of Eq. (21) has the benefit of adding few new parameters, which makes it easier to constrain them robustly. Variations of the  $(\Sigma, \mu)$  model with alternative assumptions about the time and scale dependence of deviations from GR will be explored in a follow-up paper [123].

This phenomenological approach is defined only in linear theory, while possible approaches to define similar functional forms of deviations from GR on all scales have been proposed e.g. in Ref. [124], allowing the use of halo-model based approaches as proposed in Refs. [125, 126] for  $(\Sigma, \mu)$  models. However these methods have not yet been tested for the parameterization of  $(\Sigma, \mu)$  considered here, so we restrict our analysis of DES Y3  $3 \times 2$ pt measurements to linear scales by imposing scale cuts as described in Sec. II D.

In order to model the impact of  $\Sigma_0 - \mu_0$  on the 2PCF, we modify the COSMOSIS baseline pipeline to use the Weyl potential power spectrum  $P_{\Phi\Phi}(k)$  when computing weak lensing

observables. This is in contrast to the fiducial analysis, which assumes the Poisson equation:

$$k^2\Phi = \frac{3}{2}\Omega_m H_0^2 \delta/a. \quad (22)$$

Although the impact of  $\Sigma_0$  can be computed simply modifying the lensing kernel used for 2PCF computations in Eq. (3) (as was done in DES-Y1Ext), we choose to use the Weyl potential directly as it facilitates more flexible applications to other parameterizations of modified gravity and new physics affecting growth, as used in e.g. Refs. [123, 127].

To model  $3\times 2$ pt observables, we need both the Weyl potential auto-correlation  $P_{\Phi\Phi}(k, z)$  and its correlation with the matter density  $P_{\Phi\delta}(k, z)$ . We compute their linear predictions using MGCAMB v3.0<sup>12</sup> [128], modifying its interface with COSMOSIS. The corresponding non-linear spectra are then obtained using a non-linear scaling factor:

$$P_{\Phi\Phi}^{\text{NL}}(k, z) = \frac{P_{\delta\delta}^{\text{NL}}(k, z)}{P_{\delta\delta}^{\text{L}}(k, z)} P_{\Phi\Phi}^{\text{L}}(k, z), \quad (23)$$

where the NL and L superscripts refer respectively to the HALOFIT non-linear and linear predictions of  $P(k, z)$  and we use the same non-linear boost to get the cross-power spectrum  $P_{\Phi\delta}^{\text{NL}}(k, z)$ .

We modify the  $\Lambda$ CDM modeling pipeline so that power spectra of fields derived from the Weyl gravitational potential, namely the convergence  $\kappa$  and magnification, are computed directly using the projected Weyl potential auto- and cross-power spectra. The angular power spectra  $C(\ell)$  are computed using a version of Eq. (3) with  $k^4 P_{\Phi\Phi}(k, z)$  replacing  $P_{\delta\delta}(k, z)$  for  $C_{\kappa\kappa}(\ell)$ . Similarly  $k^2 P_{\Phi\delta}(k, z)$  replaces  $P_{\delta\delta}(k, z)$  for  $C_{\kappa\delta_g}(\ell)$ . In this formulation, the lensing kernel from Eq. (4) instead reads:

$$q_{\kappa}^i(\chi) = \chi \int_{\chi}^{\chi_h} d\chi' \left( \frac{\chi' - \chi}{\chi'} \right) n_{\kappa}^i(z(\chi')) \frac{dz}{d\chi'}. \quad (24)$$

Appropriate adjustments must also be made for the modeling of galaxy clustering to account for contributions from magnification as shown in Eq. (5). Additionally, we compute the GI NLA intrinsic alignment contributions by modifying Eq. (6) such that:

$$P_{\text{GI}}(k, z) = A_1(z) k^2 P_{\Phi\delta}(k, z), \quad (25)$$

used to compute  $C_{\kappa I}(\ell)$  using the lensing kernel in Eq. (24). In a fully rigorous treatment, the modified Newtonian potential  $\Psi$  should determine the alignments of galaxies' intrinsic shapes. However, we choose to model the tidal alignment contributions (corresponding to the I term in the GI and II power spectra of Eqs. (6) and (25)) using the matter power spectrum  $P_{\delta\delta}(k, z)$  modified by  $\mu$ , by neglecting the impact of anisotropic stress. The angular power spectra  $C(\ell)$  of Eq. (5)

computed with the Weyl gravitational potential are then converted into real-space 2PCFs  $\xi_{\pm}(\theta)$ ,  $\gamma_t(\theta)$ , and  $w(\theta)$  following the same procedure as in  $\Lambda$ CDM.

We checked that this modified COSMOSIS pipeline reproduces  $\Lambda$ CDM results, with negligible shifts in parameter estimation, as shown in Appendix B. We note that the matter power spectrum  $P_{\delta\delta}(k, z)$  computed by MGCAMB shows an unexpected dependence on  $\Sigma_0$  at large scales, for  $k < 10^{-2} \text{ Mpc}/h^{-1}$ . This dependence leads in turn to a slight dependence of the clustering 2PCF  $w(\theta)$  on  $\Sigma_0$  for  $\theta$  above 100 arcmin, more significantly for the highest redshift bins. Its impact on  $\Sigma_0$  constraints is however negligible at DES Y3  $3\times 2$ pt sensitivity, with a change in the posterior value computed with simulated clustering measurements alone of 0.3% for  $\Sigma_0 = 1.5$  compared to GR.

The CAMB `dverk` routine fails due to MGCAMB implementation of the evolution of perturbations, for a large set of  $\Sigma_0 - \mu_0$  values satisfying

$$\mu_0 > 2\Sigma_0 + 1. \quad (26)$$

We thus impose a prior excluding this region of parameter space.

As opposed to other cosmological models, we will not test for consistency of  $\Sigma_0 - \mu_0$  results against an alternative IA model such as TATT. Although its use has not been fully validated against simulations for instance in non-GR theories, the NLA model allows for amplitude and redshift dependences and propagation of  $\Sigma_0 - \mu_0$  modification of gravity to IA as described above. We therefore adopt the NLA model as in other models and as in previous  $\Sigma_0 - \mu_0$  studies such as DES-Y1Ext,[120, 121]. However, to get the next-order terms, the perturbative derivation of the TATT model in [129] assumes GR and would need to be re-derived to capture the tidal physics in modified gravity theories with a similar level of generality. Adopting the TATT model for the  $\Sigma_0 - \mu_0$  analysis would amount to using a GR IA model which would accurately describe IA in the case of  $\Sigma_0 - \mu_0$  consistent with 0 but could potentially bias results if not. Additionally, we do not make use of the Fast-PT algorithm to predict the IA and galaxy bias models, and instead use the simple linear galaxy bias (as validated in [71]) and the IA NLA model using the matter and Weyl power spectra directly as computed by MGCAMB. We note that we do not use the linear alignment (LA) model, in which case the IA signal is sourced by the linear Weyl and matter power spectra: indeed, although non-linear scales are removed through the scale cuts procedure described in Section IID, we model non-linearities using HALOFIT so that non-linear information left over after scale cuts is modeled.

## F. Binned $\sigma_8(z)$

Finally, we test  $\Lambda$ CDM predictions of the evolution of structure growth without assuming a particular physical mechanism by using what we will refer to as the binned  $\sigma_8(z)$  model. This model introduces amplitudes  $A_i^{P_{\text{lin}}}$  which scale the linear matter power spectrum in redshift bins  $i$  associated

<sup>12</sup> <https://github.com/sfu-cosmo/MGCAMB>.

with our lens galaxy sample. For our fiducial MagLim lens sample, the edges of the redshift ranges used to define tomographic bins are  $[0, 0.4, 0.55, 0.7, 1.5]$ . In other words, when we perform our model calculations, in the range  $z \in [0, 0.4)$  we multiply the linear matter power spectrum by  $A_1^{P_{\text{lin}}}$ , in the range  $z \in [0.4, 0.55)$  we multiply  $P_{\text{lin}}(k)$  by  $A_2^{P_{\text{lin}}}$ , and so on. As a model-agnostic test of  $\Lambda$ CDM growth history, this is in a sense a successor to the growth-geometry split analysis of Ref. [130].

Because it introduces step functions in the linear growth factor, this parameterization implies delta function features in the linear growth rate  $f(z)$ . These spikes have no impact on the external RSD modeling because none of the  $f\sigma_8$  measurements in that likelihood fall on our  $z$ -bin boundaries. We neglect their effect on RSD contributions DES galaxy clustering.<sup>13</sup>

In practice we fix  $A_1^{P_{\text{lin}}} = 1$  and sample over  $A_{2-4}^{P_{\text{lin}}}$ . Because our measurements are sensitive to the products  $A_s A_i^{P_{\text{lin}}}$ , where  $A_s$  is the primordial power spectrum amplitude, if we varied all four  $A_i^{P_{\text{lin}}}$  amplitudes the parameters would be completely degenerate with  $A_s$ . By fixing<sup>14</sup>  $A_1^{P_{\text{lin}}}$ , we thus avoid those degeneracies and our parameterization of binned  $\sigma_8(z)$  reads:

$$\Theta_{\sigma_8(z)} \in \{A_1^{P_{\text{lin}}} \equiv 1, A_2^{P_{\text{lin}}}, A_3^{P_{\text{lin}}}, A_4^{P_{\text{lin}}}\}. \quad (27)$$

This parameterization lends itself to a physical interpretation:  $A_s$  controls the amplitude of structure observed in the lowest redshift lens bin, while the binned  $\sigma_8(z)$  amplitudes provide a consistency test of whether the time-evolution of the growth of structure is consistent with the  $\Lambda$ CDM prediction

$$A_2^{P_{\text{lin}}} = A_3^{P_{\text{lin}}} = A_4^{P_{\text{lin}}} = 1 \quad (\Lambda\text{CDM}). \quad (28)$$

We will additionally report constraints on a set of derived parameters,<sup>15</sup>

$$\sigma_8^{[\text{bin } i]} \equiv \sigma_8 \sqrt{A_i^{P_{\text{lin}}}}, \quad (29)$$

which correspond to the value of  $\sigma_8$  expected at redshift  $z = 0$  based on the amplitude of structure in redshift bin  $i$ .

When we include both *Planck* and low-redshift measurements of structures (from DES or external RSD data), we treat

the CMB measurements as an additional high- $z$  bin, and introduce an additional amplitude  $A_{\text{CMB}}^{P_{\text{lin}}}$ . In practice, we implement this by passing the product  $A_{\text{CMB}}^{P_{\text{lin}}} A_s$  as the  $A_s$  input to CAMB when computing CMB observables. To be fully self-consistent, the amount of that lensing smoothing of the CMB power spectra should account for the modulation of the line-of-sight matter power spectrum by the  $A_i^{P_{\text{lin}}}$  parameters. For simplicity, we have chosen not to model this. Instead, when we include CMB constraints for the binned  $\sigma_8(z)$  model, we marginalize over the lensing smoothing amplitude  $A_L$  [131] (matching the parameter used in *Planck* analyses) in order to remove late-time LSS information from the CMB likelihood. We neglect the dependence of the Integrated Sachs-Wolfe effect on late-time modified growth, as sensitivity to the effect is limited by cosmic variance.

Note that the phenomenology probed by this  $A_i^{P_{\text{lin}}}$  parameterization is to similar to a  $\Sigma_0 - \mu_0$  description of modified gravity with fixed  $\Sigma = 0$  and  $\mu(z)$  defined as a piecewise function of  $z$ . It is therefore comparable to models studied in e.g. Refs. [122, 132, 133]. The distinction between our binned  $\sigma_8(z)$  model and modified gravity parameterizations is largely one of interpretation rather than modeling specifics. Here we are framing binned  $\sigma_8(z)$  as a consistency test of  $\Lambda$ CDM rather than a physical model, so we use the same modeling choices as in  $\Lambda/w$ CDM, including fiducial scale cuts and HALOFIT as our model for the nonlinear matter power spectrum, in contrast to the more conservative approach adopted in the  $\Sigma_0 - \mu_0$  model above.

In our binned  $\sigma_8(z)$  analysis we include the shear ratio likelihood when analyzing DES  $3 \times 2$ pt. The shear ratio was forecasted to strengthen the constraints on the binned  $\sigma_8(z)$  amplitudes relative to  $3 \times 2$ pt alone. Because shear ratio measurements probe the relative distances between a given lens bin and different source bins, they isolate geometric information and will be insensitive to the  $A_i^{P_{\text{lin}}}$  parameters. This means shear ratio measurements help break degeneracies between the binned  $\sigma_8(z)$  amplitudes, nuisance parameters related to photometric redshift uncertainties, intrinsic alignments, and cosmological parameters which affect both geometry and growth (namely  $\Omega_m$ ). As we are treating binned  $\sigma_8(z)$  as a consistency test of  $\Lambda$ CDM rather than a physical model, we argue that we can include it without additional validation of the small-scale modeling.

#### IV. EXTERNAL DATA

We consider the DES  $3 \times 2$ pt likelihood described above in comparison to and in combination with measurements from other cosmological experiments. We use the same external measurements as in DES-Y3KP, using public likelihoods from most constraining datasets available at the time of this analysis. These include, as summarized in Table III:

- cosmic microwave background (CMB) temperature and polarization anisotropies measurements by the *Planck* satellite as described in Sec. IV A,
- distances and growth from 6dFGS, MGS, eBOSS DR16 baryon acoustic oscillations (BAO) and redshift space

<sup>13</sup> A fully consistent calculation here would account for the enhanced RSD contributions to  $w(\theta)$ . This effect could in principle be used to empirically constrain the smoothness of the linear growth factor's redshift evolution, but we neglect it for simplicity because its impact on overall constraining power is likely to be small, and implementing it would require significant updates to our modeling pipeline.

<sup>14</sup> The choice to fix the amplitude for the lowest redshift  $i = 1$  as opposed to some other bin was arbitrary. While a different choice would affect the inferred values of the  $A_i^{P_{\text{lin}}}$  parameters, the resulting model would have the same degrees of freedom, so would not affect the physical interpretation of the results — i.e. the inferred  $\sigma_8$  values.

<sup>15</sup> Note that the decision to include constraints on the derived  $\sigma_8^{[\text{bin } i]}$  parameters as part of the presentation of our binned  $\sigma_8(z)$  model results was made after unblinding.

Observables	Data
CMB ( <i>Planck</i> in text)	<i>Planck</i> 2018 TTTEEE-lowE (no lensing)
BAO	eBOSS DR16: LRGs, ELGs, QSOs, Lyman- $\alpha$ QSOs + 6dFGS + MGS
RSD	eBOSS DR16: LRGs, ELGs, QSOs + MGS
SN	Pantheon sample (2018)

TABLE III. External data used as measurements of additional observables: cosmic microwave background (CMB), baryon acoustic oscillations (BAO), redshift-space distortions (RSD), supernovae (SN).

distortions (RSD) data as described in Sec. IV B,

- supernovae (SN) distance modulus from Pantheon as described in Sec. IV C.

When performing combined analyses of these probes, we assume they are uncorrelated (except for BAO and RSD, whose correlations are taken into account in published likelihoods) so we simply multiply their likelihoods.

### A. Cosmic microwave background

The cosmic microwave background temperature and polarisation primary anisotropies carry information about density and tensor perturbations at the time of the last scattering surface. In addition, effects such as reionization and gravitational lensing caused by large scale structures produce secondary anisotropies carrying information about the evolution of the Universe since the CMB emission. In recent decades, measurements of CMB anisotropies have led to the most powerful existing constraints on  $\Lambda$ CDM cosmological parameters.

In this paper we therefore use the *Planck* 2018 TTTEEE-lowE likelihood presented in Ref. [11]. This likelihood combines three components: a high- $\ell$  likelihood based on measurements of multipoles  $30 \leq \ell \leq 2508$  for the temperature (TT) angular power spectrum and  $30 \leq \ell \leq 1996$  for the TE and EE spectra (*plik*), and two low- $\ell$  likelihoods of the temperature TT (*commander*) and the polarization EE (*simall*) spectra on multipoles  $2 \leq \ell \leq 29$ . To facilitate the study of how cosmological model extensions impact the offset between  $S_8$  constraints from  $3 \times 2$ pt and CMB temperature and polarization, we do not include the CMB lensing likelihood.

When analyzing data we use the full *Planck* likelihood provided as part of the COSMOSIS standard library. This full likelihood includes 47 nuisance parameters where 21 parameters are marginalized over, 13 of which have Gaussian priors provided with the public *Planck* likelihood. In the case of the  $\Sigma_0 - \mu_0$  model, in order to limit computing time we instead use the *Planck plik-lite* likelihood, which includes the effects of *Planck* nuisance parameter marginalization and only requires us to sample the absolute calibration parameter  $A_p$ . We have checked that it gives equivalent results to using the full likelihood in this extended model.

For simulated analyses of DES  $3 \times 2$ pt combined with external data, we use a simplified *Planck* likelihood based on Ref. [134] using two Gaussian likelihoods for  $\ell < 30$  and

using the TTTEEE *Planck plik-lite* likelihood for  $\ell \geq 30$ . We also replace the power spectra measurements with model predictions at our fiducial cosmological parameters. For both simulated and real analyses, when we include CMB measurements we marginalize over the optical depth to recombination  $\tau$  with a flat prior in the range  $[0.01, 0.8]$ .

### B. Baryon acoustic oscillations and redshift space distortions

Baryon acoustic oscillations in the early Universe imprinted features in the matter distribution at a characteristic scale which can be detected as an excess of galaxy pairs separated by a certain distance in the local universe. By measuring the relationship between redshift and the angular diameter distance associated with that excess, we can use BAO as a standard ruler to constrain the expansion history of the Universe.

We use the combinations of BAO likelihoods from eBOSS Data Release (DR) 16 [135] to provide measurements of the Hubble parameter  $H(z_i)$  and the evolution of the comoving angular distance  $d_A(z_i)$ . More specifically, we use likelihoods from the reanalysis of BOSS DR 12 Luminous Red Galaxies (LRGs) (dropping its highest redshift measurements), eBOSS LRGs, Emission Line Galaxies (ELGs), quasars (QSOs) and Lyman- $\alpha$  QSOs. These BAO measurements are made at effective redshifts of  $z_{\text{eff}} = 0.38, 0.51, 0.70, 0.84, 1.48, 2.33$ , respectively. We additionally include BAO measurements from two lower signal-to-noise galaxy samples, 6dFGS [136] and MGS [137]. These likelihoods are based on measurements of the quantity  $d_V(z) \equiv [cz(1+z)^2 d_A^2(z)/H(z)]^{1/3}$  evaluated at effective redshifts of  $z_{\text{eff}} = 0.11$  for 6dFGS and  $z_{\text{eff}} = 0.15$  for MGS.

As an external constraint on the growth rate of structure, we include the eBOSS DR16 redshift-space distortions measurements. RSD likelihoods include constraints on the growth of cosmic structure via constraints on the quantity  $f(z_i)\sigma_8(z_i)$ , where  $f$  is the linear growth rate. We use the eBOSS DR16 RSD measurements including a reanalysis of BOSS DR12 RSD measurements, LRGs, ELGs and QSOs, at the same redshifts as BAO measurements. We also use the MGS RSD measurement from [138] at  $z_{\text{eff}} = 0.15$ . When both BAO and RSD measurements from a given sample are included, we account for their covariance using the public eBOSS DR16 likelihoods.

It is worth noting that the RSD likelihoods we use are in the form of marginalized constraints on the quantity  $f\sigma_8$  at sample redshifts, and that those constraints are derived quantities from analyses which assume a  $\Lambda$ CDM template for RSD features in the galaxy distribution. When studying models beyond- $\Lambda$ CDM, care must be taken in using these likelihoods, as it is possible that inaccuracies in that template could lead to biases in beyond- $\Lambda$ CDM cosmological parameter inferences. Studies of this in e.g. Ref. [139] demonstrated that using GR-based templates they were able to obtain unbiased  $f\sigma_8$  constraints for modified gravity simulations, as long as the modified gravity model did not induce scale-dependent structure growth modifications. Given this, we follow the final eBOSS cosmology analysis [15], which uses these same RSD mea-

measurements to constrain  $w$ CDM,  $\Omega_k$ ,  $w_0 - w_a$ ,  $\Sigma_0 - \mu_0$ , and massive neutrino cosmologies, and proceed with including RSD measurements among our external likelihoods. Given the use of these measurements to constrain neutrino mass (e.g. in [15]), we assume that they are likely also safe to use for our  $N_{\text{eff}} - m_{\text{eff}}$  model, but we highlight that this assumption may be worth investigating for future, more precise, analyses.

### C. Supernovae

Type Ia supernovae are a key cosmological probe that was originally used to discover the accelerated expansion of the universe. Here we adopt the Pantheon SN Ia sample [8], which combines objects detected and followed up by several different surveys (Pan-STARRS, Sloan Digital Sky Survey (SDSS), Supernova Legacy Survey (SNLS)). The resulting sample contains 1048 SN Ia spanning the redshift range  $0.01 < z < 2.26$ .

The Pantheon likelihood assigns a Gaussian likelihood to the measured SN distance moduli,  $\mu = 5 \ln [d_L/10\text{pc}]$ . It provides a full covariance of these measurements, accounting for cosmic variance and the impact of measurement systematics. We model the distance modulus as

$$\mu = m_B - M + \alpha x_1 - \beta \mathcal{C} + \Delta_M + \Delta_B. \quad (30)$$

Here,  $x_1$  and  $\mathcal{C}$  are the light curve width and color respectively,  $\Delta_M$  is a distance correction based on the host-galaxy mass of the SN, and  $\Delta_B$  is a distance correction based on predicted biases from simulations. The calibration parameters  $\alpha$ ,  $\beta$ ,  $\Delta_M$  are fit to data as described in Ref. [8], while  $\Delta_B$  is calibrated using simulations. The absolute magnitude  $M$  is a nuisance parameter that we marginalize over in our analysis with a flat prior  $-20 < M < -18$ .

## V. ANALYSIS PROCEDURE AND VALIDATION

Our analysis procedure can be divided into five stages. These steps, described in more detail below, proceed as follows:

1. We analyze a fiducial synthetic data vector — that is, we analyze a noiseless model prediction at a fiducial set of  $\Lambda$ CDM parameters as if it were data. In this paper, the term ‘simulated analysis’ will refer to this kind of analysis of synthetic data. (Sec. V A)
2. We validate scale cuts and modeling choices by analyzing a series of alternative simulated data vectors that have been ‘contaminated’ by systematics or by modeling choices which are more complex than those used in our baseline model. (Sec. V B)
3. We perform a set of analysis tests using the fiducial synthetic data vector to study the robustness of our results against changes in the model used for parameter estimation. (Sec. V C)

4. We repeat the previous step’s robustness tests against variations of the model for real data, without unblinding the cosmology results. Once we completed these robustness tests, a draft of this paper and the analysis plan documented in it underwent a stage of DES internal review (Sec. V C).
5. Finally, we reveal our cosmology results, assess tension metrics, compute model comparison metrics, and describe the results in Sec. VI.

This procedure is designed to ensure as far as possible that decisions on how to structure the analysis are not influenced by knowledge of how they affect the main results. Since the data we are working with have already been unblinded for the  $\Lambda/w$ CDM models in DES-Y3KP, we opted not to use the two-point-function transformation blinding method [140]. Instead, we simply blinded at the parameter level, post-processing chains to shift marginalized posterior means onto fiducial values via unknown offsets. Additionally, until pre-unblinding internal review was completed, we avoided looking at comparisons between theory and model predictions for observables, tension metrics between datasets, and any model comparison statistics between extended cosmological models and  $\Lambda$ CDM.

Several analysis choices used in this work were adjusted after  $\Lambda$ CDM unblinding, in line with changes made to the analysis in DES-Y3KP. These changes are: the choice of the MagLim rather than the REDMAGIC [65, 141, 142] lens sample, and the removal of the two highest redshift MagLim lens bins (which would have been bins 5 and 6) from the fiducial analysis. As noted in Sec. II C, we have also adopted the simpler two-parameter NLA description of intrinsic alignments instead of the five-parameter TATT model used in DES-Y3KP. We made the choice to use NLA largely because the DES-Y3KP  $\Lambda$ CDM analysis did not find that the TATT model is favored over NLA. We emphasize that while these aspects of our analysis have been shaped by findings for unblinded  $\Lambda$ CDM results, these changes were frozen in before analyzing any data for the beyond- $\Lambda$ CDM models.

### A. Fiducial synthetic data

We initially validate our analysis by performing a series of analyses of synthetic measurements — that is, analyses of the data vector predicted from our fiducial model, with no noise. With the exception of the change in data vector, this analysis is identical to our eventual analysis of real data: the synthetic DES  $3 \times 2$ pt data are generated using the same redshift distributions used for the final analysis, and the likelihood is evaluated using a covariance produced at our fiducial cosmology using the same analytical calculations described in Ref. [74]. In addition to synthetic DES  $3 \times 2$ pt measurements, we additionally produce synthetic versions of the external likelihoods for simulated combined analyses.

We begin with a baseline simulated analysis: using our fiducial model, we analyze synthetic data vectors produced using those same calculations. This can be thought of as a ‘best case

scenario’ where our model calculations are exactly correct so that we can estimate the expected constraining power and the relationship between marginalized posteriors and parameters’ input values.

In some cases when constraints are weak, prior volume effects cause marginalized confidence intervals for parameters to be offset from their input ( $\Lambda$ CDM) values. This occurs because the prior in our full parameter space can be highly non-uniform when projected onto certain cosmological variables. This occurs notably for the synthetic *Planck*-only results, which prefer  $\Omega_k < 0$  at  $1.4\sigma$ . This offset from flatness, which can be understood in terms of the CMB’s well-known geometric degeneracy [143], is in the same direction as what has been previously reported for the analysis of real CMB data but at a lower significance. The preference for  $\Omega_k < 0$  goes away when the *Planck* likelihood is combined with low-redshift geometric likelihoods.

We also see offsets in the DES  $3\times 2$ pt and  $3\times 2$ pt+BAO+RSD+SN  $N_{\text{eff}} - m_{\text{eff}}$  constraints, which is due to a positive degeneracy between  $N_{\text{eff}}$  and  $m_{\text{eff}}$  constraints for the lower redshift probes, as well as degeneracies between both of those parameters and  $H_0$ . Adding CMB information introduces a powerful constraint on  $N_{\text{eff}}$ , breaking those degeneracies and causing the marginalized posterior distribution for the All-data constraints to be more reflective of the input parameter values. Given this concern about the projection effects, for  $N_{\text{eff}} - m_{\text{eff}}$  we will focus primarily on constraints from DES Y3  $3\times 2$ pt and all external data (*i.e.* BAO+RSD+SN+*Planck*) rather than DES alone.

In the  $\Sigma_0 - \mu_0$  model,  $\mu_0$  measurements by DES Y3  $3\times 2$ pt alone are prior-dominated so will not be reported. We additionally note that a  $\Sigma_0 - S_8$  degeneracy causes a slight offset for the DES-only  $\Sigma_0$  posterior, though the resulting constraints are consistent with the input value. The addition of external RSD or *Planck* data enable precise measurements of  $\mu_0$ , in turn leading to more precise constraints on  $\Sigma_0$ .

## B. ‘Contaminated’ synthetic data

Next, we analyze synthetic  $3\times 2$ pt data that have been ‘contaminated’ by various effects. The goal here is to test robustness of our results to modeling complexities and systematic effects which are not included in our fiducial model. To verify this, we compare the results of the analysis of contaminated synthetic data to those from our baseline simulated analysis, both computed in  $\Lambda$ CDM. This allows us to quantify the impact these modeling uncertainties have on parameter estimates and model comparison statistics used to evaluate tensions with  $\Lambda$ CDM. The priority here — and what we can evaluate most accurately, given the lack of in-depth study of and available modeling tools for describing systematics in extended cosmological models — is to assess whether unmodeled systematics can produce a false detection of tension with  $\Lambda$ CDM. Specifically, we study three alterations to the synthetic data:

- **Nonlinear bias + baryons:** A realization of baryonic feedback and non-linear galaxy bias is added

to the synthetic  $3\times 2$ pt observables. Baryonic feedback effects are added using the method described in Refs. [144, 145], and are based on the OWLS hydrodynamic simulations [146, 147] with large AGN feedback according to the prescription of Ref. [148]. Nonlinear galaxy bias is modeled using an effective 1-loop description with renormalized nonlinear bias parameters as in Refs. [149–152]. This synthetic data is produced with the same contaminations used in Ref. [71] to define scale cuts for DES-Y3KP. Analyzing it allows us to verify that the scale cuts defined for  $\Lambda$ CDM continue to protect against these small-angle systematics in the beyond- $\Lambda$ CDM models we consider.

- **Euclid emulator:** The HALOFIT computation for the nonlinear, gravity-only matter power spectrum is replaced with that of EUCLID EMULATOR [153]. Checking that this alternate nonlinear prescription does not shift our results is a test of robustness against inaccuracies of the small-scale power spectrum modeling.
- **Magnification  $3\sigma$  offset:** The magnification coefficients  $C_i$  (see Eq. (5)) are offset from their fiducial values by three times their uncertainty, where the latter is determined in Ref. [89] using the Balrog simulations. The coefficients used are  $\mathcal{C} = [1.97, 1.74, 2.93, 2.97]$  compared to the fiducial coefficients  $\mathcal{C} = [0.42, 0.30, 1.76, 1.94]$ . This  $3\sigma$  offset is designed to demonstrate the robustness of results to the amplitude of the magnification signal and validate the decision to fix the  $\mathcal{C}$  values at their fiducial values.

To facilitate these tests we adopt the newly-developed FASTISMORE (Fast Importance Sampling for MOdel Robustness Evaluation) scheme, which is presented in more detail in Appendix C and Ref. [154]. Briefly, the approach exploits the fact that if our analysis is robust against a given systematic, the shift in posteriors should be small when the data is ‘contaminated’ with that effect. This allows us to use the posterior from a baseline chain (run using uncontaminated synthetic  $3\times 2$ pt observables) as a proposal distribution for estimating the posterior for a contaminated datavector via importance sampling (IS). Doing this allows us to quickly verify whether the change in posterior is indeed negligible. If quality statistics indicate that the IS posterior estimate is of good quality, it can be used to quantify the effect of the systematic on parameter estimates and model comparison statistics. If the estimate is poor, this indicates we need to run a new chain to estimate the contaminated posterior.

Once we have obtained reliable posterior estimates, we assess shifts in the marginalized constraints on the beyond- $\Lambda$ CDM parameters. We quantify this following Ref. [130], defining the marginalized posterior shift  $\Delta_\theta$  for parameter  $\Theta$  as

$$\Delta_\Theta = \frac{\bar{\Theta}_A - \bar{\Theta}_B}{\sqrt{(\bar{\Theta}_A - \Theta_A^{\text{low68}})^2 + (\Theta_B^{\text{up68}} - \bar{\Theta}_B)^2}}. \quad (31)$$

Here,  $\bar{\Theta}$  is the parameter’s posterior-weighted mean, and the labels  $A$  and  $B$  correspond to the baseline and alterna-

tive (contaminated) synthetic data vectors, defined such that  $\hat{\Theta}_A > \hat{\Theta}_B$ . The parameter  $\Theta_A^{\text{low}68}$  is the lower bound of the 68% confidence interval for posterior  $A$ , while  $\Theta_B^{\text{up}68}$  is the upper bound of the 68% confidence interval for posterior  $B$ . Thus the denominator of Eq. (31) is an effective  $1\sigma$  error for parameter  $\Theta$ , accounting for possible asymmetry in the marginalized posteriors. We consider a parameter shift to be negligible if  $\Delta_\Theta < 0.3$ .

We perform these checks for DES  $3\times 2\text{pt}$  and DES  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}+\text{Planck}$  posteriors for each beyond- $\Lambda\text{CDM}$  model, as well as  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}$  (leaving out *Planck*), using simulated external data likelihoods produced at the same fiducial cosmology as the synthetic DES data. Results are shown in the “Alt data” rows of Fig. 1, with points and error bars indicating the mean and marginalized 68% confidence interval for each parameter. Where error bars are not visible they are smaller than the size of the data point. In that plot the  $N_{\text{eff}}^{(0)}$  constraints are for the model which varies  $N_{\text{eff}}$  only, while  $N_{\text{eff}}^{(m)}$  shows the  $\Delta N_{\text{eff}}$  constraint from the  $N_{\text{eff}} - m_{\text{eff}}$  model. Points with  $\Delta_\Theta > 0.3$  are highlighted.

Nearly all shifts evaluated were below the  $0.3\sigma$  threshold, meaning these beyond- $\Lambda\text{CDM}$  parameter estimates are robust against each of the considered systematics. The only exception to this occurs for the binned  $\sigma_8(z)$  model’s response to changes in the assumed magnification parameter. For the binned  $\sigma_8(z)$  DES  $3\times 2\text{pt}$  results we see  $\Delta_{A_3^{P_{\text{lin}}}} = 0.31$ , and for  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}+\text{Planck}$  binned  $\sigma_8(z)$  results we find  $\Delta_{A_3^{P_{\text{lin}}}} = 0.37$ . We note that these numbers are close to the desired threshold, especially relative to our sampling uncertainty of  $\Delta_\Theta \pm \sim 0.04$ , and so are not very concerning.

We also check whether these contaminations affect our assessment of whether an extended model is favored relative to  $\Lambda\text{CDM}$ . We do this by comparing the values of the Suspiciousness model comparison statistic  $S$ , defined in Appendix F, evaluated in our contaminated and baseline simulated analyses. For a base model  $\mathcal{M}_0$  (e.g.  $\Lambda\text{CDM}$ ) whose parameter space is a subspace of extended model  $\mathcal{M}_X$ , we define Suspiciousness so that  $\ln S < 0$  indicates a preference for  $\mathcal{M}_X$ . Of the several model comparison statistics that we will ultimately report as part of our results, we use Suspiciousness here because it is readily calculable from importance-sampled posteriors. Fig. 2 shows the changes produced by systematic contamination in  $\ln S$  relative to the expected amount of scatter, for DES  $3\times 2\text{pt}$  and DES  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}+\text{Planck}$  constraints. The largest shift occurs for  $N_{\text{eff}}$ -vs- $\Lambda\text{CDM}$ , where analyzing the EUCLID EMULATOR synthetic data shifts  $\ln S$  by  $\sim 1\sigma$  in the limit that posteriors are Gaussian. We should therefore interpret model comparison results for  $N_{\text{eff}}$  with some caution. Otherwise, all systematics considered cause negligible changes in Suspiciousness and thus are unlikely to result in a false detection of tension with  $\Lambda\text{CDM}$ .

### C. Robustness to model variations

We additionally study how parameter constraints respond to changes in our model. By comparing results obtained using

alternative modeling choices to those obtained from our fiducial model, we assess the robustness of our findings relative to those model variations. As before, we quantify this comparison in terms of the parameter shift  $\Delta_\Theta$  defined in Eq. (31), and we consider any shifts with  $\Delta_\Theta < 0.3$  to be negligible. The model changes considered are:

- **TATT** - We use a five-parameter TATT intrinsic alignment model [129] instead of the two-parameter NLA model used in the present baseline analysis. This model, which was the fiducial IA model in DES-Y3KP, has significantly more flexibility to describe IA scale and redshift dependence, allowing it to capture IA from tidal alignment and tidal torquing, as well as the response to the density-weighted tidal field. We use the same parameters and prior ranges as in DES-Y3KP. Comparison with an analysis using TATT allows us to test the robustness of our results against our choice of intrinsic alignment model. As explained in Sec. II C, for each of the beyond- $\Lambda\text{CDM}$  models we perform a pre-unblinding model comparison between TATT and NLA to test whether there is tension with the choice of NLA as the fiducial IA model for that cosmology. Note that this set of tests are not performed for  $\Sigma_0 - \mu_0$  because the TATT modeling tools are not available for our modified gravity calculations (see Sec. III E).
- **Varying  $X_{\text{Lens}}$** : We marginalize over a parameter  $X_{\text{Lens}}$  which multiplies the galaxy bias terms appearing in  $\gamma_t(\theta)$  calculations, thus allowing the galaxy-galaxy lensing observables to have a different bias parameter than galaxy clustering. Such an effect was discovered after unblinding the  $\Lambda\text{CDM}$   $3\times 2\text{pt}$  analysis using the REDMAGIC lens sample, and while it is still under investigation, it is thought to be due to an unaccounted for systematic related to lens sample selection [142]. This effect motivated the choice of MagLim over REDMAGIC as the fiducial lens sample in DES-Y3KP. While no evidence was found in  $\Lambda\text{CDM}$  for  $X_{\text{Lens}} \neq 1$  for our data vector, which uses the four-bin MagLim lens sample, we include results marginalizing over parameter to test the robustness of our beyond- $\Lambda\text{CDM}$  constraints to the presence of this kind of systematic in the real data analysis. Note that a similar effect with independent  $X_{\text{Lens}}$  values for each redshift bin is able to capture the issues with the fifth and sixth MagLim bins which led to their removal from the analysis. Studies in  $\Lambda\text{CDM}$  have shown that the first four MagLim bins we analyze are consistent with  $X_{\text{Lens}} = 1$  in this redshift-dependent formulation as well [66, 142]. Thus to limit the parameter space of these tests, here we consider only a single redshift-independent  $X_{\text{Lens}}$  parameter. As an additional exploration of this effect, in Appendix G we study the response of the extended models to synthetic data produced with  $X_{\text{Lens}} \neq 1$ .
- **Hyperrank**: Instead of using the four  $\Delta z_s$  photo- $z$  bias parameters of our fiducial model, we use the hyperrank [155] method to marginalize over uncertainties in

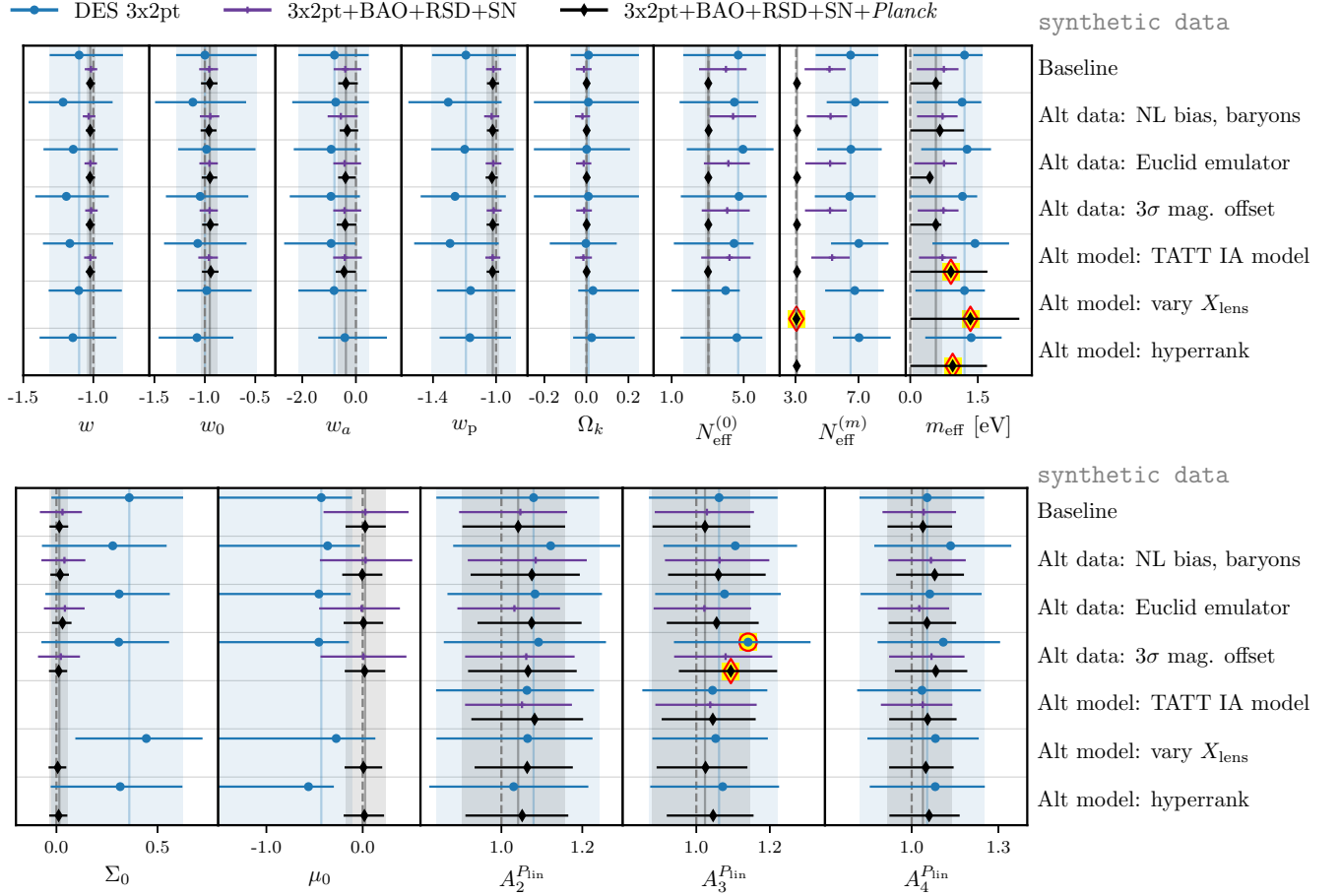


FIG. 1. Simulated analysis constraints on beyond- $\Lambda$ CDM model parameters, showing robustness against systematics and model variations. Points and error bars show the mean and 68% confidence interval for marginalized parameter constraints, and points that are offset from the baseline by more than  $0.3\sigma$ , according to Eq. (31), are highlighted in yellow and red. To facilitate comparison between rows, solid vertical lines and shaded regions show the location and 68% confidence interval of the baseline point for DES  $3\times 2$ pt and DES  $3\times 2$ pt+BAO+RSD+SN+*Planck* results in blue and black respectively. Dashed vertical lines show the  $\Lambda$ CDM values used to generate the synthetic data vectors. We use  $N_{\text{eff}}^{(0)}$  to identify the effective number of relativistic degrees of freedom when no mass is included in the model,  $N_{\text{eff}}^{(m)}$  for that parameter in the  $N_{\text{eff}} - m_{\text{eff}}$  model. For parameters where the combination of all data (DES  $3\times 2$ pt and all external data) is much more constraining than DES  $3\times 2$ pt, a version of this plot with narrower axis ranges can be found in Fig. 19 of Appendix D.

the source sample redshift distributions. This method uses three ‘rank’ parameters to sample an ensemble of 1000 realizations of source  $n(z)$  histograms that were generated using a three-step Dirichlet (3sDir) sampling method [59, 156, 157]. (Throughout this paper we will use “hyperrank” as shorthand for what might more properly be referred to as the 3sDir+hyperrank method.) Compared to the fiducial approach, hyperrank captures more information about uncertainties in the shapes of those redshift distributions as well as correlations between different source bins. Comparing constraints obtained using hyperrank to our fiducial model will indicate how sensitive our results are to our characterization of source photo- $z$  uncertainties. Note that in DES-Y3KP, it was found that using hyperrank for  $\Lambda$ CDM produced a  $0.53\sigma$  shift in  $S_8$  (see Appendix E1 of that Reference and our Appendix D 4).

Because we are using the TATT comparisons to evaluate our choice of fiducial IA model (except for the  $\Sigma_0 - \mu_0$  model), we run TATT analyses for all three principal data combinations (DES  $3\times 2$ pt alone, only low redshift probes: DES  $3\times 2$ pt+BAO+RSD+SN, and all data combined: DES  $3\times 2$ pt+BAO+RSD+SN+*Planck*). We run  $X_{\text{Lens}}$  and hyperrank analyses for DES  $3\times 2$ pt only, except for the  $\Sigma_0 - \mu_0$ , binned  $\sigma_8(z)$  and  $N_{\text{eff}} - m_{\text{eff}}$  models. This choice was motivated by non-negligible shifts seen for binned  $\sigma_8(z)$  and  $N_{\text{eff}} - m_{\text{eff}}$  in the blinded real-data tests for DES  $3\times 2$ pt-alone. Additional chains were run for  $\Sigma_0 - \mu_0$  as well because that model also primarily affects structure growth.

Results showing the impact of these model variations on simulated analyses are shown in the ‘Alt model’ lines of Fig. 1,

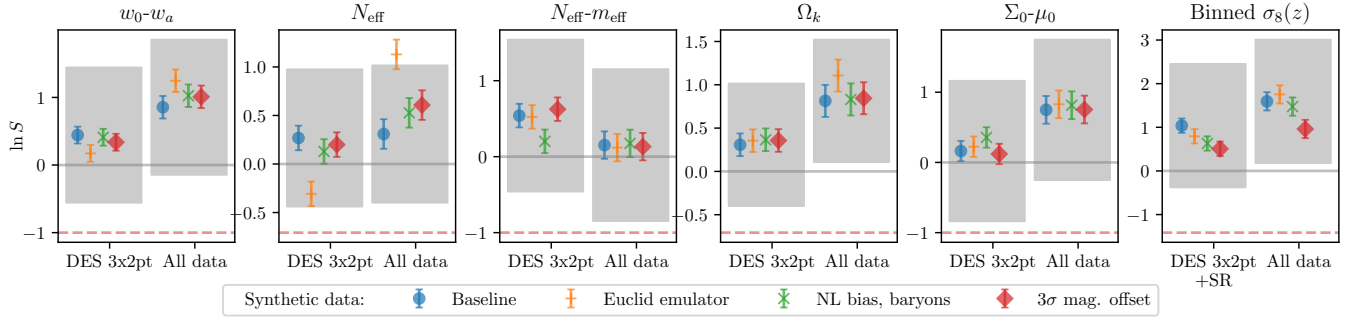


FIG. 2. Impact of contamination by unmodeled systematics on the Suspiciousness model comparison metric evaluated for synthetic  $\Lambda$ CDM data. Panels show Suspiciousness evaluated between different beyond- $\Lambda$ CDM models and  $\Lambda$ CDM, with the marker styles indicating different contaminations. Blue circular markers indicate the baseline case where the synthetic data is produced using the same model as used for parameter estimation. Studying whether contaminations shift  $\ln S$  towards more negative values than the baseline measurement tells us whether they are likely to produce a false detection of beyond- $\Lambda$ CDM physics. Shaded regions indicate the expected  $\pm 1\sigma$  scatter around the baseline for Gaussian posteriors, and the red dashed line denotes the value of  $\ln S$  that would produce a  $1\sigma$  preference for the beyond- $\Lambda$ CDM model (though as we discuss in Appendix F  $\ln S$  is expected to follow a  $\chi^2$  distribution, such that the true uncertainty is more skewed toward negative values of  $\ln S$  than the shaded regions, making the test conservative). Error bars on points indicate sampling uncertainty. Results are shown for the analysis of DES  $3 \times 2$ pt alone and  $3 \times 2$ pt+BAO+RSD+SN+*Planck* (All data).

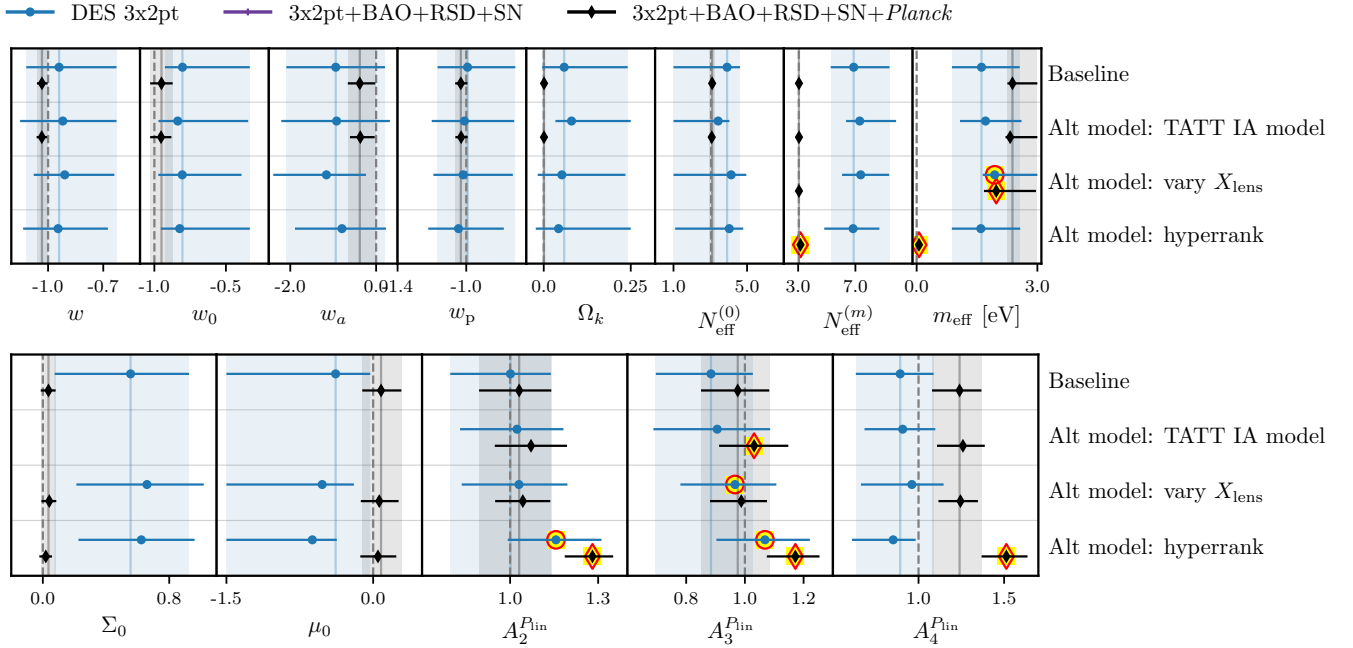


FIG. 3. Beyond- $\Lambda$ CDM parameters constraints for real data and the impact of model variations. Points and error bars show the mean and 68% confidence interval for marginalized constraints on extended model parameters, and shifts larger than  $0.3\sigma$  according to Eq. (31) are highlighted in yellow and red. Dashed gray lines show parameter values that correspond to  $\Lambda$ CDM. Solid vertical lines and shaded regions mark the DES  $3 \times 2$ pt and combination of all data, DES  $3 \times 2$ pt+BAO+RSD+SN+*Planck* baseline results, in blue and black respectively, to facilitate comparison between rows. No points are shown in second row of  $\Sigma_0$  and  $\mu_0$  constraints because of the lack of TATT IA model implementation within the modified gravity pipeline. A version of this plot with narrower axis ranges can be found in Fig. 20 of Appendix D 1.

and results for real data are shown in Fig. 3.<sup>16</sup> For synthetic and real data, other than for  $N_{\text{eff}} - m_{\text{eff}}$  and binned  $\sigma_8(z)$ ,

<sup>16</sup> Before unblinding cosmological parameter estimates or model comparisons with  $\Lambda$ CDM, at this stage we studied only differences between the

alternative and baseline chain constraints, working with a version of Fig. 3 where baseline means were subtracted from all numbers, causing all the points in the top row to be on zero. The Figure was updated after unblinding to show the actual parameter values of the constraints.

all shifts due to these model variations were below  $0.3\sigma$ . The non-negligible parameter shifts occurring for  $N_{\text{eff}} - m_{\text{eff}}$  and binned  $\sigma_8(z)$  have motivated adjustments to analyses choices for those models. Further detailed discussion of those shifts can be found for  $N_{\text{eff}} - m_{\text{eff}}$  in Appendix D 2 and for binned  $\sigma_8(z)$  in Appendix D 3, respectively, which we briefly summarize here. First, the sensitivity of the  $N_{\text{eff}} - m_{\text{eff}}$  constraints to model variations seem to be caused by prior volume effects associated with an unconstrained part of parameter space at small  $\Delta N_{\text{eff}}$ . Since adjusting our prior to remove that part of parameter space restores robustness, this motivates our choice to report  $N_{\text{eff}} - m_{\text{eff}}$  results using priors that require either  $\Delta N_{\text{eff}} > 0.047$  or  $m_{\text{th}} < 10$  eV, as noted at the end of Sec. III D.

Second, for binned  $\sigma_8(z)$  the most concerning parameter shifts occur for hyperrank, which causes the  $A_i^{P_{\text{lin}}}$  amplitudes to change by  $\sim 1 - 2\sigma$  relative to their baseline model estimates. Given this, we report binned  $\sigma_8(z)$  results for both the baseline and hyperrank versions of the analysis, using their comparison as a rough estimate for the impact of systematic uncertainties related photometric redshift estimates. We note that model variations most severely impact inferences about  $\sigma_8$  in the lowest redshift bin, which then propagates to affect  $A_i^{P_{\text{lin}}}$  parameters because they are defined relative to bin 1. Importantly, the derived parameters  $\sigma_8^{[\text{bin } i]} \equiv \sigma_8[A_i^{P_{\text{lin}}}]^{1/2}$ , which are more closely related to the observed amplitude of LSS, are more robust to modeling variations especially when we combine DES  $3 \times 2\text{pt}$  with external data.

## VI. RESULTS

We now show the principal results of our analysis. This section is organized as follows: In Table IV, we show a summary of marginalized constraints on individual parameters. In Secs. VIA-VIF, we report and discuss constraints for each of the cosmological models studied in this work. In Secs. VIG and VIH we examine how these cosmological models and other model variations impact estimates of  $S_8$  and  $\sum m_\nu$  respectively. While discussions of results for individual models touch on tensions and model comparisons, Secs. VII and VIJ, present more details about the definitions and determination of tension and model comparison statistics, respectively.

In this section, parameter estimates shown with two-sided error bars report the mean and 68% confidence interval of marginalized one-dimensional posteriors. One-sided errors report 95% confidence bounds. Before unblinding, we decided that we would report results from combined datasets only if the  $p$ -value associated with the Suspiciousness tension metric, described in Appendix E and Ref. [158], is greater than 0.01. We evaluate that metric for two data combinations: DES  $3 \times 2\text{pt}$  versus the external low-redshift data combination BAO+RSD+SN, and *Planck* versus the combination of all non-CMB data,  $3 \times 2\text{pt}$ +BAO+RSD+SN.

		DES 3x2pt	All External	All data
$w\text{CDM}$	$w_0$	$-0.94^{+0.31}_{-0.18}$	$-1.04^{+0.03}_{-0.03}$	$-1.03^{+0.03}_{-0.03}$
$w_0 - w_a$	$w_0$	$\geq -1.40$	$-0.94^{+0.08}_{-0.08}$	$-0.95^{+0.08}_{-0.08}$
	$w_a$	$-0.94^{+1.15}_{-1.15}$	$-0.45^{+0.36}_{-0.28}$	$-0.38^{+0.36}_{-0.28}$
	$w_p$	$-0.99^{+0.28}_{-0.17}$	$-1.04^{+0.04}_{-0.03}$	$-1.03^{+0.04}_{-0.03}$
$\Omega_k$	$10^2 \Omega_k$	$\geq -16$	$0.08^{+0.18}_{-0.18}$	$0.09^{+0.17}_{-0.17}$
$N_{\text{eff}}$	$N_{\text{eff}}$	$\leq 7.84$	$3.10^{+0.16}_{-0.17}$	$3.10^{+0.15}_{-0.16}$
$\Delta N_{\text{eff}} > 0.047$	$\Delta N_{\text{eff}}$	-	$\leq 0.36$	$\leq 0.34$
	$m_{\text{eff}} [\text{eV}]$	-	$\leq 0.18$	$\leq 0.14$
$m_{\text{th}} < 10 \text{eV}$	$\Delta N_{\text{eff}}$	-	$\leq 0.23$	$\leq 0.28$
	$m_{\text{eff}} [\text{eV}]$	-	$\leq 0.42$	$\leq 0.20$
$\Sigma_0 - \mu_0$	$\Sigma_0$	$0.56^{+0.37}_{-0.48}$	$0.37^{+0.12}_{-0.09}$	$0.04^{+0.05}_{-0.05}$
	$\mu_0$	-	$0.20^{+0.22}_{-0.22}$	$0.08^{+0.21}_{-0.19}$
Binned $\sigma_8(z)$	$A_2^{P_{\text{lin}}}$	$1.00^{+0.14}_{-0.21}$	$0.92^{+0.14}_{-0.23}$	$1.03^{+0.11}_{-0.14}$
	$A_3^{P_{\text{lin}}}$	$0.88^{+0.14}_{-0.19}$	$0.95^{+0.17}_{-0.30}$	$0.98^{+0.11}_{-0.13}$
	$A_4^{P_{\text{lin}}}$	$0.89^{+0.20}_{-0.26}$	$1.22^{+0.17}_{-0.33}$	$1.24^{+0.13}_{-0.16}$
	$A_{\text{CMB}}^{P_{\text{lin}}}$	-	$0.89^{+0.10}_{-0.22}$	$1.04^{+0.04}_{-0.06}$
	$\sigma_8^{[\text{bin } 1]}$	$0.75^{+0.05}_{-0.05}$	$0.83^{+0.09}_{-0.06}$	$0.78^{+0.02}_{-0.02}$
	$\sigma_8^{[\text{bin } 2]}$	$0.74^{+0.06}_{-0.07}$	$0.79^{+0.06}_{-0.06}$	$0.79^{+0.04}_{-0.04}$
	$\sigma_8^{[\text{bin } 3]}$	$0.70^{+0.06}_{-0.07}$	$0.80^{+0.07}_{-0.07}$	$0.76^{+0.04}_{-0.04}$
	$\sigma_8^{[\text{bin } 4]}$	$0.70^{+0.10}_{-0.09}$	$0.90^{+0.05}_{-0.04}$	$0.86^{+0.04}_{-0.05}$
	$\sigma_8^{[\text{CMB}]}$	-	$0.78^{+0.02}_{-0.02}$	$0.79^{+0.01}_{-0.01}$
Binned $\sigma_8(z)$ , hyperrank	$A_2^{P_{\text{lin}}}$	$1.16^{+0.16}_{-0.16}$	$0.92^{+0.14}_{-0.23}$	$1.28^{+0.07}_{-0.09}$
	$A_3^{P_{\text{lin}}}$	$1.07^{+0.15}_{-0.17}$	$0.95^{+0.17}_{-0.30}$	$1.17^{+0.08}_{-0.10}$
	$A_4^{P_{\text{lin}}}$	$0.85^{+0.13}_{-0.24}$	$1.22^{+0.17}_{-0.33}$	$1.51^{+0.12}_{-0.14}$
	$A_{\text{CMB}}^{P_{\text{lin}}}$	-	$0.89^{+0.10}_{-0.22}$	$1.26^{+0.03}_{-0.04}$
	$\sigma_8^{[\text{bin } 1]}$	$0.73^{+0.04}_{-0.04}$	$0.83^{+0.09}_{-0.06}$	$0.72^{+0.01}_{-0.01}$
	$\sigma_8^{[\text{bin } 2]}$	$0.78^{+0.06}_{-0.06}$	$0.79^{+0.06}_{-0.06}$	$0.81^{+0.02}_{-0.02}$
	$\sigma_8^{[\text{bin } 3]}$	$0.75^{+0.07}_{-0.06}$	$0.80^{+0.07}_{-0.07}$	$0.77^{+0.03}_{-0.03}$
	$\sigma_8^{[\text{bin } 4]}$	$0.67^{+0.07}_{-0.09}$	$0.90^{+0.05}_{-0.04}$	$0.88^{+0.04}_{-0.04}$
	$\sigma_8^{[\text{CMB}]}$	-	$0.78^{+0.02}_{-0.02}$	$0.80^{+0.01}_{-0.01}$

TABLE IV. Marginalized constraints on beyond- $\Lambda$ CDM parameters for DES Y3  $3 \times 2\text{pt}$ , all external data (*Planck*+BAO+RSD+SN), and all data ( $3 \times 2\text{pt}$ +*Planck*+BAO+RSD+SN). Two-sided constraints report the mean and 68% confidence interval for each parameter's marginalized posterior. One-sided constraints report 95% bounds.

### A. Results: $w_0 - w_a$

We start with dynamical dark energy described by parameters  $w_0$  and  $w_a$ . The marginalized constraints are shown in

the left panel of Fig. 4. We find

$$\begin{aligned} w_0 &\geq -1.4, & w_a &= -0.9 \pm 1.2 & \text{DES Y3} & (32) \\ w_0 &= -0.95 \pm 0.08, & w_a &= -0.4^{+0.4}_{-0.3} & \text{DES Y3 + External.} \end{aligned}$$

DES Y3  $3 \times 2$ pt data alone weakly constrain<sup>17</sup>  $w_0$  and  $w_a$ , but are statistically consistent with the cosmological-constant values of  $(w_0, w_a) = (-1, 0)$ . When combined with external data, the constraints tighten considerably. The combined constraints find  $w_a < 0$  at about  $1\sigma$ , and are thus consistent with the standard model. In the  $w_0 - w_a$  plane, the DES  $3 \times 2$ pt data alone have similar constraining power to *Planck* alone. Additionally,  $3 \times 2$ pt+BAO+RSD+SN (without *Planck*) produce constraints comparable those from the combination of all data considered.

The pivot equation-of-state (see Sec. III A) is

$$\begin{aligned} w_p &= -0.99^{+0.28}_{-0.17} & \text{DES Y3} \\ &= -1.03^{+0.04}_{-0.03} & \text{DES Y3 + External,} \end{aligned} \quad (33)$$

where for the DES Y3  $3 \times 2$ pt and all-data constraints, the pivot redshifts are  $z_p = 0.24$  and  $z_p = 0.27$ , respectively. Note that DES alone does give a two-sided constraint on  $w_p$ , unlike on  $w_0$ . The right panel of Fig. 4 shows the constraints in the  $(w_p, w_a)$  plane. The DES Y3  $3 \times 2$ pt data again qualitatively agree with CMB alone, but are somewhat more centered on the  $\Lambda$ CDM model values of  $w_p = -1$ ,  $w_a = 0$ . We note that the  $3 \times 2$ pt  $w_a$  constraints are slightly tighter than the latest Pantheon+ supernovae measurements [159], which report a 68% confidence interval width of 2.8 on  $w_a$ , compared to 2.3 from DES Y3  $3 \times 2$ pt.

Fig 5 presents a more detailed view of the most powerful constraints in the  $w_0 - w_a$  plane, additionally showing constraints from the external low redshift data (BAO+RSD+SN) alone and *Planck* combined with low-redshift geometric data (BAO+SN). Looking at individual parameter constraints, we note that the marginalized posterior mean for  $w_0$  is essentially the same for all data combinations considered here, and that compared to constraints from BAO+RSD+SN alone, adding  $3 \times 2$ pt data causes the  $w_a$  estimate to shift downwards by about  $1\sigma$ . Adding *Planck* constraints to that moves the all-data  $w_a$  estimate slightly lower, but not by a significant amount, and produces constraints that are very similar to the *Planck*+BAO+SN data combination. All constraints are statistically consistent with  $\Lambda$ CDM parameter values.

Overall, the DES  $3 \times 2$ pt data strengthen the case that the  $(w_0, w_a)$  model parameters are in excellent agreement with the  $\Lambda$ CDM values  $w_0 = -1$ ,  $w_a = 0$ . We discuss the comparison between the present results and DES Y1  $3 \times 2$ pt constraints on  $w_0 - w_a$  [50] in Appendix H.

## B. Results: $\Omega_k$

Fig. 6 shows constraints on curvature in the  $\Omega_k - \Omega_m$  plane. We see that curvature is not strongly constrained by the DES Y3  $3 \times 2$ pt data alone, and that the constraints on curvature from all data (DES  $3 \times 2$ pt+BAO+RSD+SN+*Planck*) are identical to those from *Planck* combined with low-redshift geometric probes (BAO+SN). The DES  $3 \times 2$ pt data do contribute constraining power when combined with other low redshift data by breaking degeneracies in the full parameter space. Specifically, when combined with BAO+RSD+SN, DES  $3 \times 2$ pt data lower the upper bound on  $\Omega_k$  by constraining  $\Omega_m$  and thus helping to break a degeneracy between  $\Omega_m$  and  $\Omega_k$ . This decreases the width of the marginalized 68% confidence interval on  $\Omega_k$  by 20%.

We recover the well-documented (see e.g. [11, 160–166]) finding that constraints from *Planck* alone favor negative  $\Omega_k$  at roughly three sigma and are significantly offset from the low-redshift constraints. That offset is along the direction of the primary CMB anisotropies’ geometric degeneracy [143], indicated in Fig. 6 with a gray dashed line<sup>18</sup>, and when that degeneracy is broken with low-redshift observables the constraints shift to become consistent with  $\Omega_k = 0$ . Quantifying the tension between the *Planck*-only and  $3 \times 2$ pt+BAO+RSD+SN posteriors, we find the Suspiciousness  $p$ -value to be exactly at our threshold of 0.01 for combining data. Given this, we report constraints for the combination of all the data, but additionally report the constraints from  $3 \times 2$ pt+BAO+RSD+SN (without *Planck*). The marginalized constraints on  $\Omega_k$  are

$$\begin{aligned} \Omega_k &\geq -0.16 & \text{DES Y3} \\ &= 0.001 \pm 0.002 & \text{DES Y3 + External} & (34) \\ &= -0.03^{+0.04}_{-0.03} & \text{DES Y3 + External, no CMB} \end{aligned}$$

Compared to the combination of all external data, the inclusion of  $3 \times 2$ pt in the DES Y3 + External constraints listed above narrows the 68% confidence interval range by only 6%. Compared to the eBOSS analysis of Ref. [167], in which  $\Omega_k$  is measured from the combination of CMB, BAO, and SN, our constraints are about 10% weaker, likely due to differences in analysis procedures (for example, we vary neutrino mass, while that analysis fixes it). Our “no CMB” constraints are slightly tighter than those reported in Ref. [168] from a different combination of low-redshift probes.

## C. Results: $N_{\text{eff}}$

Much like curvature, the number of relativistic species  $N_{\text{eff}}$  is weakly constrained by DES  $3 \times 2$ pt alone, but combining the

<sup>17</sup> We report two-sided  $3 \times 2$ pt-only constraints on  $w_a$  because its one-dimensional marginalized posterior is not bounded by the  $w_a$  priors, but note that the constraint is strongly influenced by the intersection of the posterior with the upper  $w_0$  prior boundary.

<sup>18</sup> The line is drawn for constant shift parameter  $\mathcal{R} \propto \sqrt{\Omega_m} h^2 (1 + z_*) d_A(z_*)$  corresponding to the Planck+BAO constraints reported in Ref. [11], where  $z_*$  is the redshift of recombination.

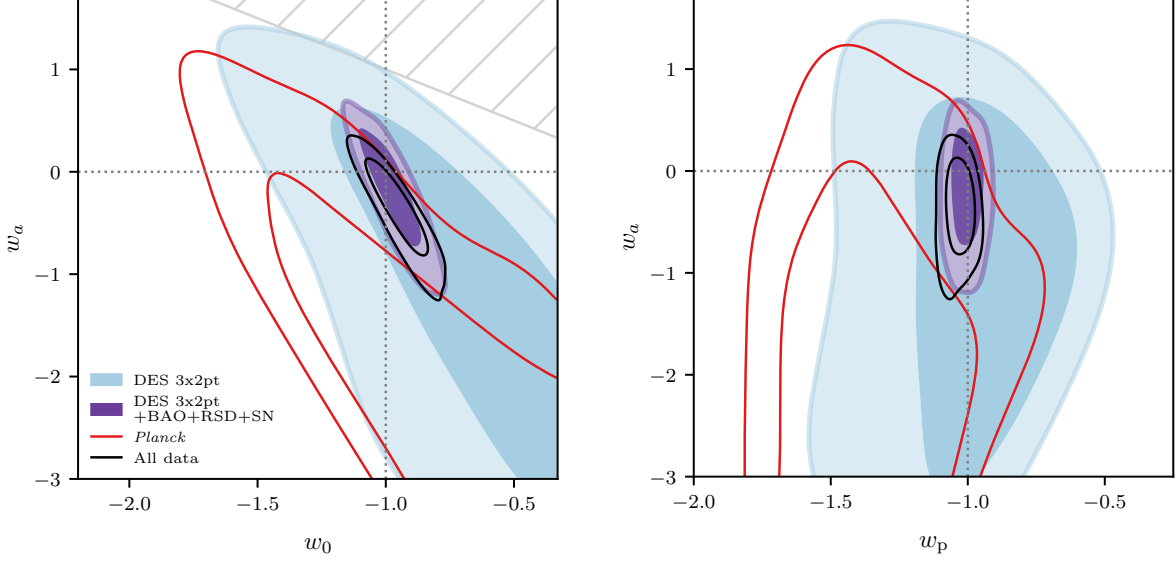


FIG. 4. Constraints on the dynamical dark-energy  $w_0 - w_a$  model. The left panel shows the constraints on  $w_0$  and  $w_a$ , while the right panel shows constraints on the derived pivot value  $w_p$  and again  $w_a$ . Contours show 68% and 95% confidence regions. The pale blue contours are DES  $3 \times 2$ pt, the purple contours is the low- $z$  combination of  $3 \times 2$ pt+BAO+RSD+SN, the open red contours are for *Planck*, and the open black contours represent everything combined. The pivot redshifts derived for the  $w_p - w_a$  are 0.24 for DES  $3 \times 2$ pt only, 0.21 for  $3 \times 2$ pt+BAO+RSD+SN, and 0.27 for all data constraints. The gray hatched region shows the part of parameter space removed by the prior that requires  $w_0 + w_a < 0$ .

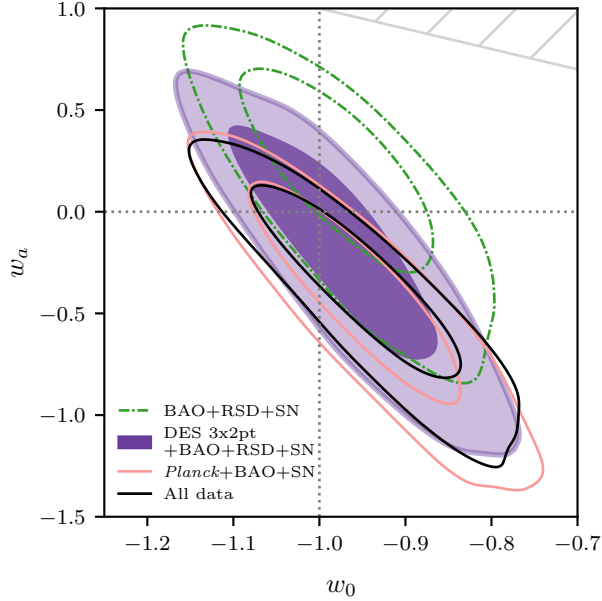


FIG. 5. A more detailed look at how different data contribute constraining power to  $w_0$  and  $w_a$ . The purple contours showing  $3 \times 2$ pt+BAO+RSD+SN and black contours showing all data constraints are the same as those in Fig. 4. The green contours show constraints from the external low- $z$  data alone (BAO+RSD+SN), while the pink contours show constraints from *Planck* combined with low redshift geometric probes only (BAO+SN). The gray hatched region shows the excluded region where  $w_0 + w_a > 0$ .

DES with external data leads to modest improvements. The constraints are

$$\begin{aligned} N_{\text{eff}} &\leq 7.8 && \text{DES Y3} \\ &= 3.10^{+0.15}_{-0.16} && \text{DES Y3 + External.} \end{aligned} \quad (35)$$

Constraints on the number of relativistic species are given in Fig. 7. The DES  $3 \times 2$ pt and BAO+RSD+SN constraints both individually peak around  $N_{\text{eff}} \simeq 3$ , though we caution that these constraints are mainly shaped by prior projection effects. Both the  $3 \times 2$ pt and BAO+RSD+SN posteriors are unconstrained along  $N_{\text{eff}} - \Omega_b$  and  $N_{\text{eff}} - h$  degeneracy directions. Given our choice of priors, the upper bounds on  $N_{\text{eff}}$  are shaped by the upper prior bound on  $h$ , while the lower bound on  $N_{\text{eff}}$  are determined by the lower prior bound for  $\Omega_b$ . Both DES  $3 \times 2$ pt and RSD are sensitive to the amplitude of the power spectrum, which is affected by  $N_{\text{eff}}$  through changes in the redshift of matter-radiation equality, while BAO is additionally sensitive to a small phase shift caused by  $N_{\text{eff}}$ 's impact on the Silk damping scale. These probes' posteriors have different degeneracies between  $A_s$  and  $N_{\text{eff}}$ , and the overlap between them rules out small  $N_{\text{eff}}$  values, while the upper bound is still primarily determined by where the posterior intersects the  $\Omega_b$  prior. The constraints are still consistent with the standard-model value of  $N_{\text{eff}}$ . Once the  $3 \times 2$ pt+BAO+RSD+SN data are combined with *Planck*, the overall combined posterior shifts slightly compared to the *Planck*-only constraints, but remain fully consistent with  $N_{\text{eff}} \simeq 3$ .

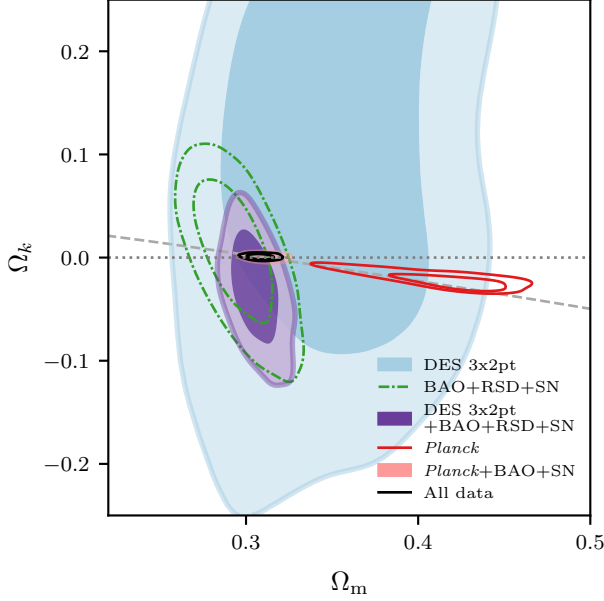


FIG. 6. Constraints on the curvature density  $\Omega_k$  and matter density  $\Omega_m$ . The diagonal dashed gray line shows the direction of the primary CMB anisotropies' geometric degeneracy [143].

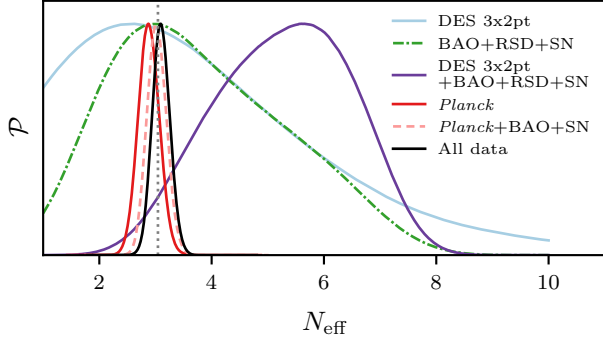


FIG. 7. Marginalized posterior on the number of relativistic species  $N_{\text{eff}}$ . The vertical dotted gray lines shows the standard model value of  $N_{\text{eff}} = 3.044$ .

#### D. Results: $N_{\text{eff}} - m_{\text{eff}}$

We next consider the  $N_{\text{eff}} - m_{\text{eff}}$  model. Recall, following discussion in Sec. V C, that this particular set of constraints is prone to parameter-space projection effects in the small- $\Delta N_{\text{eff}}$  region, so our fiducial analysis imposes the constraint  $\Delta N_{\text{eff}} > 0.047$ , where  $\Delta N_{\text{eff}} \equiv N_{\text{eff}} - 3.044$ . Making that assumption, and reporting the constraints only for all data combined, we find

$$\Delta N_{\text{eff}} < 0.34, \quad m_{\text{eff}} < 0.14 \text{ eV} \quad \text{DES Y3 + External.} \quad (36)$$

The constraints are shown in Fig. 8.

We also consider constraints for a prior requiring  $\Delta N_{\text{eff}} >$

0, but also  $m_{\text{th}} < 10 \text{ eV}$ . For this prior, the constraints from all data combined are

$$\Delta N_{\text{eff}} < 0.28, \quad m_{\text{eff}} < 0.20 \text{ eV} \quad \text{DES Y3 + External.} \quad (37)$$

This prior on the sterile neutrino's thermal mass matches the *Planck* 2018 analysis [11], which used the same CMB likelihood as us, plus *Planck* lensing constraints (which we do not include) and BOSS DR12 BAO to constrain  $\Delta N_{\text{eff}} < 0.23$  and  $m_{\text{eff}} < 0.65 \text{ eV}$ . Thus, while we find slightly weaker constraints on  $N_{\text{eff}}$ , the inclusion of DES  $3 \times 2$ pt and RSD effective measurements of the amplitude of structure at low redshifts allows us to tighten the upper bound on  $m_{\text{eff}}$  by about a factor of three.

Another interesting, if less direct, comparison can be made to the analysis of Ref. [?]. That work analyzed the *Planck* data including lensing, the full-shape BOSS DR12 measurement, and CFHTLenS weak lensing measurements at fixed  $\Delta N_{\text{eff}} = 0.047$ , to find the constraint  $m_{\text{eff}} < 1.6 \text{ eV}$ . They found that the weak lensing measurements were crucial to obtaining a tight constraint on  $m_{\text{eff}}$ . While our results are not directly comparable (our  $m_{\text{eff}}$  constraints would likely weaken if we performed our analysis at fixed, small  $\Delta N_{\text{eff}}$ ), our findings lend further support to the idea that precise cosmic shear measurements of LSS can powerfully constrain the presence of light but massive relic particles produced in the early Universe.

#### E. Results: $\Sigma_0 - \mu_0$

Next, we show results of tests of gravity on cosmological scales parametrized by  $\Sigma_0 - \mu_0$  in Fig. 9. As discussed in Sec. V A,  $\mu_0$  is not constrained by DES Y3  $3 \times 2$ pt alone so will not be reported. The Eq. (26) prior set on  $\Sigma_0 - \mu_0$  is represented by a hatched area in the figure. While the *Planck*-only contours are visibly offset from the  $3 \times 2$ pt + BAO+RSD+SN combination, the Suspiciousness tension metric comparing the two posteriors has  $p = 0.02$  (above the 0.01 threshold), so we proceed with reporting results from the combination of all data. These constraints are

$$\Sigma_0 = 0.6 \pm 0.4 \quad \text{DES Y3}$$

$$\Sigma_0 = 0.04 \pm 0.05, \quad \mu_0 = 0.08^{+0.21}_{-0.19} \quad \text{DES Y3 + External,} \quad (38)$$

the latter of which can be compared to the external-only constraint, which is  $\Sigma_0 = 0.37^{+0.12}_{-0.09}$ ,  $\mu_0 = 0.20 \pm 0.22$ . Thus the addition of DES  $3 \times 2$ pt data both tightens the constraints on  $(\Sigma_0, \mu_0)$  and shifts them to be more consistent with general relativity.

The constraint on  $\Sigma_0$  from DES Y3  $3 \times 2$ pt data alone is limited by the linear scale cuts used for this model, which actually results in the  $3 \times 2$ pt-only bounds on  $\Sigma_0$  presented here to be weaker than the comparable DES-Y1Ext result by 40%. To understand this, recall from Sec. II D that we define these linear scale cuts by iteratively removing small-angle measurements until the difference between  $3 \times 2$ pt model predictions

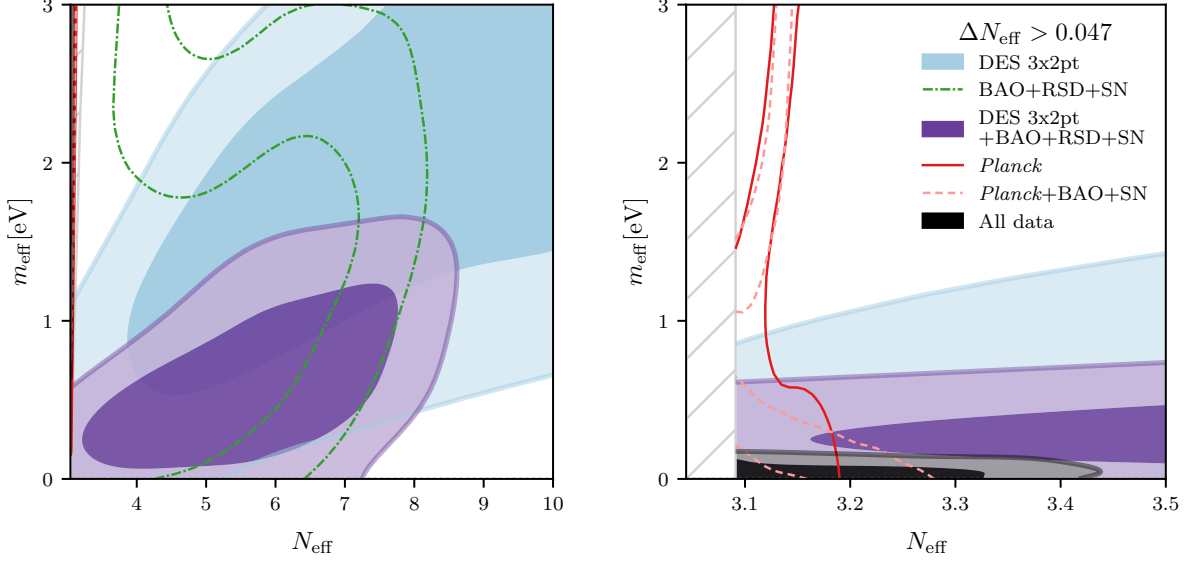


FIG. 8. Constraints on the beyond- $\Lambda$ CDM parameters of the  $N_{\text{eff}} - m_{\text{eff}}$  model. The left plot shows results obtained using the fiducial prior  $3.044 < N_{\text{eff}} < 10$ , with axes spanning the full prior range. In it, the *Planck* and all-data posteriors reside entirely inside the  $m_{\text{th}} > 10\text{eV}$  region where sterile neutrinos behave like CDM, shown in a very narrow gray hatched wedge along the left axis of the plot. The other contours show constraints from DES  $3 \times 2\text{pt}$  (blue), external low redshift data from BAO+RSD+SN (dashed green), and  $3 \times 2\text{pt}$  + BAO+RSD+SN (purple). The right panel shows chains run using an alternative prior where  $\Delta N_{\text{eff}} > 0.047$ , and with the plot range reduced to more clearly show the all-data ( $3 \times 2\text{pt}$ +BAO+RSD+SN+*Planck*) results. Because the DES  $3 \times 2\text{pt}$  and DES  $3 \times 2\text{pt}$ +BAO+RSD+SN chains have very few samples in the excluded region, the alternative prior has no effect on the purple and blue contours.

with and without nonlinear modeling are deemed insignificant. We assess the significance of that difference relative to the data covariance, which means that as measurements get more precise this method produces more stringent cuts. Quantitatively,  $\Sigma_0 - \mu_0$  constraints in the present analysis are only based on 55% of the fiducial data vector, while in DES-Y1Ext the linear cuts retained 73% of the data points. The weakened  $\Sigma_0$  constraints imply that this method for protecting against the Y3 measurements' greater sensitivity to non-linear effects reduces the S/N available for cosmology inference and indicates the need for a more sophisticated method of accommodating for nonlinear modeling uncertainties as data get more precise.

Given the offset of the *Planck* contour as well as complementary of growth measurements from RSD, which primarily constrain  $\mu_0$ , and  $3 \times 2\text{pt}$ , which constrain  $\Sigma_0$ , it is interesting to report modified gravity constraints from the combination of only low-redshift probes. This result is shown in purple in Figs. 9, leading to measurements of  $\Sigma_0 - \mu_0$  independent of CMB observables:

$$\Sigma_0 = -0.06^{+0.09}_{-0.10}, \quad \mu_0 = -0.4 \pm 0.4 \quad (39)$$

DES Y3 + External, no CMB.

Results from the combined constraints is partly driven by the BAO+RSD+SN measurement of  $\mu_0 = -0.5^{+0.4}_{-0.5}$ , indicating  $\mu_0 < 0$  at  $1\sigma$  significance. This is in contrast to the results reported in Ref. [167], in which analysis of the same BAO+RSD data assuming a fixed background cosmology produces  $\mu_0$  constraints centered on zero.

To aid in the interpretation of these results, Fig. 10 shows how different data combinations break degeneracies between the modified gravity parameters and  $S_8$ . The fact that lensing observables are sensitive to the product  $S_8 \Sigma_0$  limits the ability of DES  $3 \times 2\text{pt}$  alone to constrain  $\Sigma_0$ . By constraining  $S_8$  and  $\mu_0$ , RSD measurements of the growth rate of structure are able to break that degeneracy and thus improve constraints on  $\Sigma_0$ .

The behavior of the *Planck*-only contours can be understood in terms of two relevant degeneracies. First, *Planck*'s constraints on  $\Sigma_0$  mainly come from the impact of lensing, which smooths the high- $\ell$  peaks of the CMB power spectra. (The integrated Sachs-Wolfe effect [169, 170] also introduces some sensitivity to  $\Sigma_0$ , but cosmic variance affecting the low multipoles where that occurs limits its constraining power.) That lensing signal, like DES cosmic shear, is sensitive to the product  $S_8 \Sigma_0$ , and thus leads  $S_8$ - $\Sigma_0$  degeneracy parallel to that seen for  $3 \times 2\text{pt}$ . Second,  $\mu_0$  adds a degree of freedom to the relationship between CMB constraints on the primordial power spectrum amplitude  $A_s$  and the amplitude of density fluctuations in the late Universe. This leads to a positive degeneracy between  $S_8$  and  $\mu_0$ : for a given value of  $A_s$ , which is tightly constrained by *Planck*, larger  $\mu_0$  will enhance structure growth and thus increase  $S_8$ . Combining *Planck* with the BAO+RSD+SN data allows the RSD observables' direct measurement of  $S_8$  break both of these degeneracies.

The fact that the *Planck* contours, either from *Planck* alone or in combination with the BAO+RSD+SN low redshift data, are offset towards higher  $\Sigma_0$  than  $3 \times 2\text{pt}$  follows a trend seen

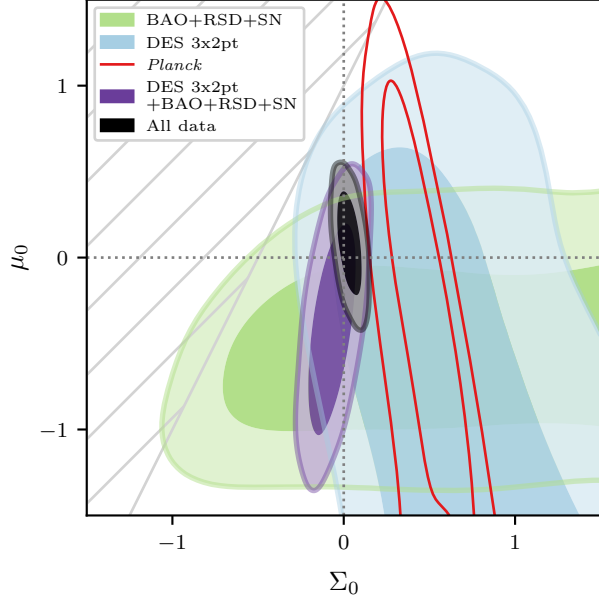


FIG. 9. Constraints on the  $\Sigma_0 - \mu_0$  modified gravity parameters, with axis ranges reflecting the parameters' prior ranges. The gray hatched region shows where samples are excluded by the  $\mu_0 < 2\Sigma_0 + 1$  requirement of MGCAMB: any overlap of contours with that region is a reflection of KDE smoothing done by GETDIST.

for DES Y1 data in both DES-Y1Ext and Ref. [11]. As noted in Ref. [11], this preference is driven by the excess smoothing of high- $\ell$  *Planck* measurements that are captured by the phenomenological  $A_L$  parameter. These are the same features that pull the *Planck*-only  $\Omega_k$  constraints towards negative values. When all data are analyzed together, the  $\Sigma_0$  constraints are in agreement with those from  $3\times 2$ pt, with the CMB measurements contributing to tightening constraints primarily by breaking the RSD posterior's weak degeneracy between  $\sigma_8$  and  $\mu_0$ .

### F. Results: binned $\sigma_8(z)$

Finally, we report constraints on the binned  $\sigma_8(z)$  model. We begin by examining the set of derived parameters  $\sigma_8^{[\text{bin } i]}$  (see Eqs. (27-29)), which correspond to the values of  $\sigma_8$  inferred from LSS observed in redshift bin  $i$ , and which we showed in Sec. V C are more robust to model variations than the sampled  $A_i^{P_{\text{lin}}}$  parameters. These constraints are:

$$\begin{aligned} \sigma_8^{[\text{bin } 1]} &= 0.75^{+0.05}_{-0.05}, \\ \sigma_8^{[\text{bin } 2]} &= 0.74^{+0.06}_{-0.07}, \\ \sigma_8^{[\text{bin } 3]} &= 0.70^{+0.06}_{-0.07}, \\ \sigma_8^{[\text{bin } 4]} &= 0.70^{+0.10}_{-0.09}, \end{aligned} \quad \text{DES Y3} \quad (40)$$

and

$$\begin{aligned} \sigma_8^{[\text{bin } 1]} &= 0.78^{+0.02}_{-0.02}, \\ \sigma_8^{[\text{bin } 2]} &= 0.79^{+0.04}_{-0.04}, \\ \sigma_8^{[\text{bin } 3]} &= 0.76^{+0.04}_{-0.04}, \\ \sigma_8^{[\text{bin } 4]} &= 0.86^{+0.04}_{-0.05}, \\ \sigma_8^{[\text{CMB}]} &= 0.792^{+0.015}_{-0.010}, \end{aligned} \quad \text{DES Y3 + External} \quad (41)$$

Fig. 11 presents these constraints in comparison to  $\Lambda$ CDM constraints on  $\sigma_8$ . In that figure, the set of lighter, unfilled data points show how the  $\sigma_8^{[\text{bin } i]}$  constraints change when use the alternative hyperrank method of marginalizing over source galaxy photo- $z$  uncertainties. We find that hyperrank induces non-negligible but still small ( $\sim 0.5\sigma$ ) shifts in  $\sigma_8^{[\text{bin } i]}$  for the  $3\times 2$ pt-only  $i \in \{2, 3\}$  measurements, and all-data  $i \in \{2, \text{CMB}\}$  measurements, while a much larger, almost  $3\sigma$  shift occurs for the all-data constraint on  $\sigma_8^{[\text{bin } 1]}$ . As is discussed in more detail in Appendices D 3 and D 4, the lack of robustness of the lowest redshift is likely due to an interaction between the source  $n(z)$  and IA modeling which is most significant at low redshifts. In the same Figure, we report additional results to facilitate interpretation of how different structure growth observables contribute to constraints. Namely, we show the combination of DES data with only geometric external data ( $3\times 2$ pt+BAO+SN) shifts constraints to slightly higher  $S_8$  in both  $\Lambda$ CDM and binned  $\sigma_8$ , but not as much as the  $3\times 2$ pt+BAO+RSD+SN data combination. Thus the combined analysis' shift towards higher  $\sigma_8$ , especially in the highest redshift bin 4, seems to be primarily driven by the RSD likelihood.

Fig. 12 translates these results to the inferred growth function  $\sigma_8(z)$ . That figure compares marginalized constraints when we vary the binned  $\sigma_8(z)$  amplitude parameters shown with data points at a few example redshifts, to the 68% confidence bands obtained from  $\Lambda$ CDM fits to DES  $3\times 2$ pt and all data ( $3\times 2$ pt+BAO+RSD+SN+*Planck*).

All measurements are within approximately  $1\sigma$  of the  $\Lambda$ CDM  $\sigma_8$  estimate. The fact that the DES-only constraints on  $\sigma_8(z)$  are consistently lower than *Planck* and that our combined constraints find  $\sigma_8^{[\text{bin } 4]}$  to be higher than  $\sigma_8^{[\text{bin } i]}$  in the other bins agrees with similar features seen in Refs. [171–173]. In those works, analyses of DESI galaxies cross correlated with *Planck* lensing, eBOSS QSO clustering, and both of those observables combined with DES Y1  $3\times 2$ pt measurements, respectively, suggest that the amplitude of structure at  $z \sim 0.8$  may be slightly higher compared to lower redshift measurements, and thus hinting at a slower growth rate than expected in  $\Lambda$ CDM. However the trends seen in these references, as well as that in our work, are only significant at the  $\sim 1\sigma$  level and thus not strong enough to motivate any kind of firm conclusion. We also note that a similar trend is not found in Ref. [122]'s binned modified gravity study using DES Y1 and BOSS DR12 LSS data. It will be interesting to monitor how new and more precise data constrain the time-evolution

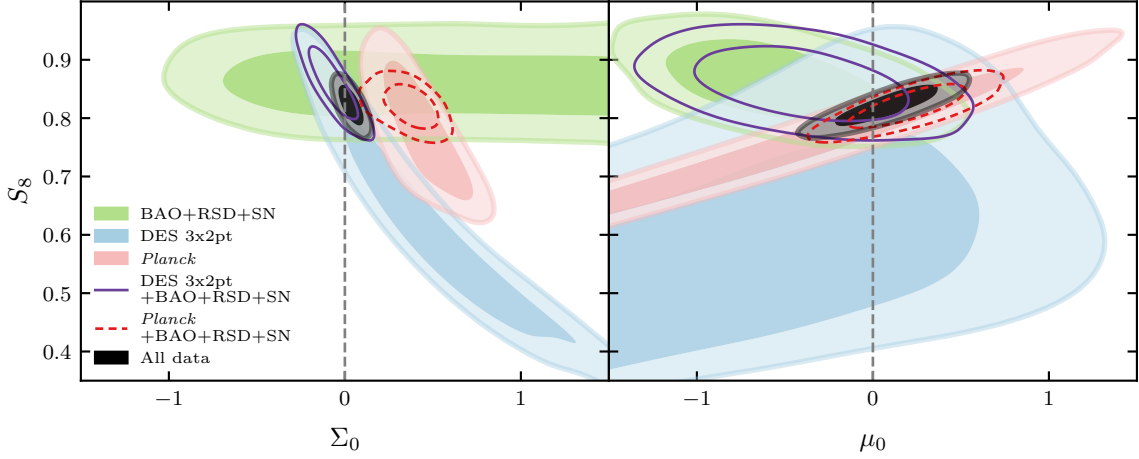


FIG. 10. Examination of how RSD growth information, as part of the BAO+RSD+SN external low redshift data combination, breaks degeneracies between  $S_8$  and modified gravity parameters  $\Sigma_0$  and  $\mu_0$  when combined with either DES 3 $\times$ 2pt or *Planck*.

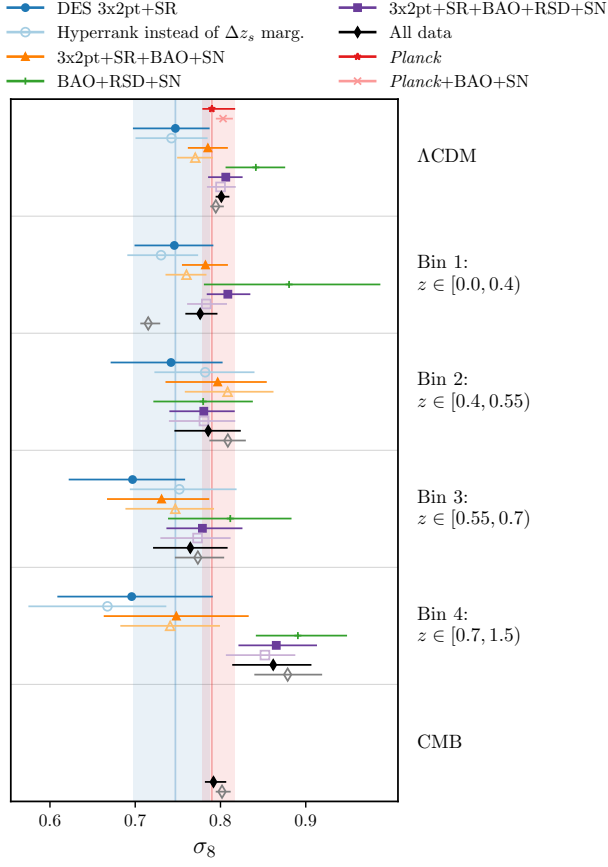


FIG. 11. Constraints on binned  $\sigma_8(z)$  compared to  $\Lambda$ CDM constraints in the top row. Points and error bars show the marginalized posterior mean and 68% confidence intervals. Unfilled, lighter markers correspond to the same data as the darker points with matching colors and shapes, but were obtained using the hyperrank marginalization over uncertainties in source galaxy redshift distributions. Shaded bands highlight the location of the top row's 3 $\times$ 2pt and *Planck*  $\Lambda$ CDM points for comparison.

of the amplitude of density fluctuations.

For completeness, we additionally report constraints on the sampled amplitude parameters used to implement this model,  $A_2^{P_{\text{lin}}}$ ,  $A_3^{P_{\text{lin}}}$ ,  $A_4^{P_{\text{lin}}}$ , and  $A_{\text{CMB}}^{P_{\text{lin}}}$  (see Eq. (27)). We emphasize that these constraints should be interpreted with caution, because in Sec. VC we found that they are not robust to a change in how we marginalize over our source photo- $z$  uncertainties. Particularly for the all-data constraints, this occurs because these sampled amplitude parameters are defined relative to bin 1, so the sensitivity to hyperrank seen for  $\sigma_8^{[\text{bin } 1]}$  propagates to  $A_i^{P_{\text{lin}}}$  inferences for higher redshift bins. Specifically, switching from our fiducial analysis to one with the hyperrank  $n_s(z)$  marginalization scheme produces shifts between  $0.5 - 2\sigma$  shifts in all of the amplitude parameters. The hyperrank-model constraints on these parameters can be found in Table IV. The constraints on these parameters using our baseline model are

$$\begin{aligned} A_2^{P_{\text{lin}}} &= 1.00^{+0.14}_{-0.21}, \\ A_3^{P_{\text{lin}}} &= 0.88^{+0.14}_{-0.19}, \\ A_4^{P_{\text{lin}}} &= 0.90^{+0.20}_{-0.26}, \end{aligned} \quad \text{DES Y3} \quad (42)$$

and

$$\begin{aligned} A_2^{P_{\text{lin}}} &= 1.03^{+0.11}_{-0.14}, \\ A_3^{P_{\text{lin}}} &= 0.98^{+0.11}_{-0.13}, \\ A_4^{P_{\text{lin}}} &= 1.24^{+0.13}_{-0.16}, \\ A_{\text{CMB}}^{P_{\text{lin}}} &= 1.04^{+0.04}_{-0.06}. \end{aligned} \quad \text{DES Y3 + External} \quad (43)$$

### G. Impact of model and analysis choices on $S_8$

In line with studies exploring whether beyond- $\Lambda$ CDM models alleviate the tension between weak lensing and *Planck*

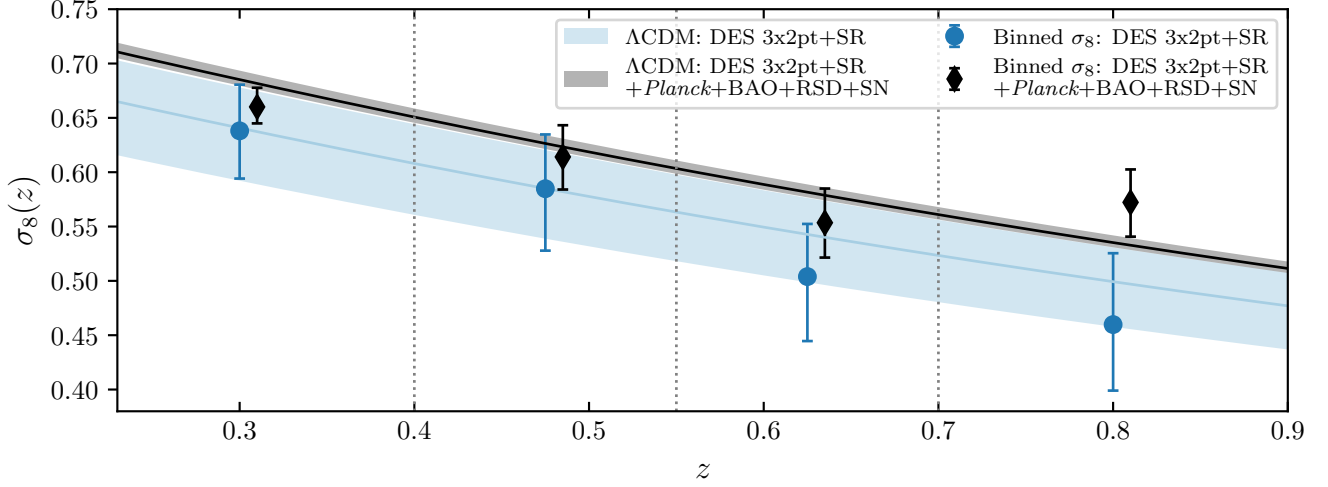


FIG. 12. Inferred  $\sigma_8(z)$  from  $\Lambda$ CDM (lines) and the binned  $\sigma_8(z)$  model (points). Lines and shaded bands show the mean and 68% confidence interval inferred from  $\Lambda$ CDM posteriors, with DES  $3\times 2$ pt + shear ratio shown in blue and the combination of all data ( $3\times 2$ pt+SR+Planck+BAO+RSD+SN) shown in black. Points with error bars show binned  $\sigma_8(z)$  results for the same data combinations, plotted at an example redshift for each lens bin. For readability, the all-data (black) points have been shifted by a small horizontal offset. Vertical dashed lines show the bin divisions used to define the binned  $\sigma_8(z)$  model's  $A_i^{P_{\text{lin}}}$  amplitudes.

$S_8$  measurements (see Sec. I and e.g. Refs. [41, 174–188]), in Fig. 13 we compare constraints on  $S_8$  obtained within a selection of cosmological models and, for comparison, analysis choices within the  $\Lambda$ CDM model. In that Figure, points show the mean and 68% confidence intervals of the marginalized  $S_8$  posterior, with different colors and marker styles corresponding to results from different sets of observables. We report constraints for DES  $3\times 2$ pt (blue), the combination of DES and other low redshift probes ( $3\times 2$ pt+BAO+RSD+SN, purple), *Planck* alone (red), *Planck* combined with low-redshift geometric probes BAO+SN to break geometric degeneracies (pink), and the combination of all data ( $3\times 2$ pt+BAO+RSD+SN+*Planck*, black). Blue and red vertical lines and bands mark the location of the baseline  $\Lambda$ CDM DES  $3\times 2$ pt and *Planck* constraints shown in the top row, to indicate the level of offset between those measurements and to facilitate comparisons with other rows.

The first group of  $S_8$  constraints shown are for beyond- $\Lambda$ CDM models. These include  $w$ CDM along with the extended models studied in this paper, except for the binned  $\sigma_8(z)$  model for which  $\sigma_8$  constraints were discussed above in Sec. VI F. In these extended models, we see that the most overlap between *Planck* and  $3\times 2$ pt  $S_8$  constraints occurs for dynamical dark energy described by  $w$ CDM and  $w_0 - w_a$ . For both  $\Lambda$ CDM and beyond- $\Lambda$ CDM models, we find that the combination of low redshift probes, i.e. combining DES Y3  $3\times 2$ pt with BAO, RSD and SN as shown in purple, measure  $S_8$  to be more consistent with *Planck* constraints than  $3\times 2$ pt alone. This repeats the same finding of DES-Y3KP (see Figs. 14 and 15 of Ref. [44]). The behavior occurs because the external geometric (BAO+SN) probes constrain  $\Omega_m$  to be at the higher end of the range allowed by the  $3\times 2$ pt-only constraints. Given the  $\Omega_m$ - $\sigma_8$  degeneracy, this leads to a

higher  $S_8$  value.

For comparison, additional blocks of points in Fig. 13 show how constraints on  $S_8$  are affected by changes to the analysis choices while retaining the  $\Lambda$ CDM model, and the impact of two extensions to  $\Lambda$ CDM which we label ad hoc models. The  $\Lambda$ CDM analysis choice variations include using the shear ratio likelihood, different scale cuts, the more general TATT IA model, the hyperrank method for marginalizing over source photo- $z$  uncertainties, and fixing neutrino mass. The ad hoc models include varying  $X_{\text{Lens}}$ , which introduces a mismatch between the galaxy bias affecting galaxy clustering and that affecting galaxy-galaxy lensing (see its description as part of the robustness tests of Sec. V C), and varying  $A_L$  [131], which scales the amount of lensing-related smoothing affecting the CMB temperature and polarization power spectra (See Sec. III F and e.g. Ref. [11]). Both these ad hoc models correspond to the introduction of a parameter to explain features in the DES Y3  $3\times 2$ pt and *Planck* data respectively, and not new physics as opposed to the beyond- $\Lambda$ CDM models considered in this analysis. While  $X_{\text{Lens}}$  has little effect on the  $3\times 2$ pt constraints and thus on the  $3\times 2$ pt-*Planck* comparison, varying  $A_L$  leads to more consistent estimates of  $S_8$  across all probe combinations shown (see also [179, 182]).

#### H. Impact of model choice on $\sum m_\nu$

Next, we examine constraints on the sum of neutrino masses  $\sum m_\nu$ . While DES  $3\times 2$ pt data are not particularly sensitive to this parameter, the fact that we vary  $\Omega_\nu h^2$  as part of our baseline analysis allows us to study how its bounds are impacted by the assumed cosmological model. With this aim, Table V reports the 95% upper bound on  $\sum m_\nu$  from

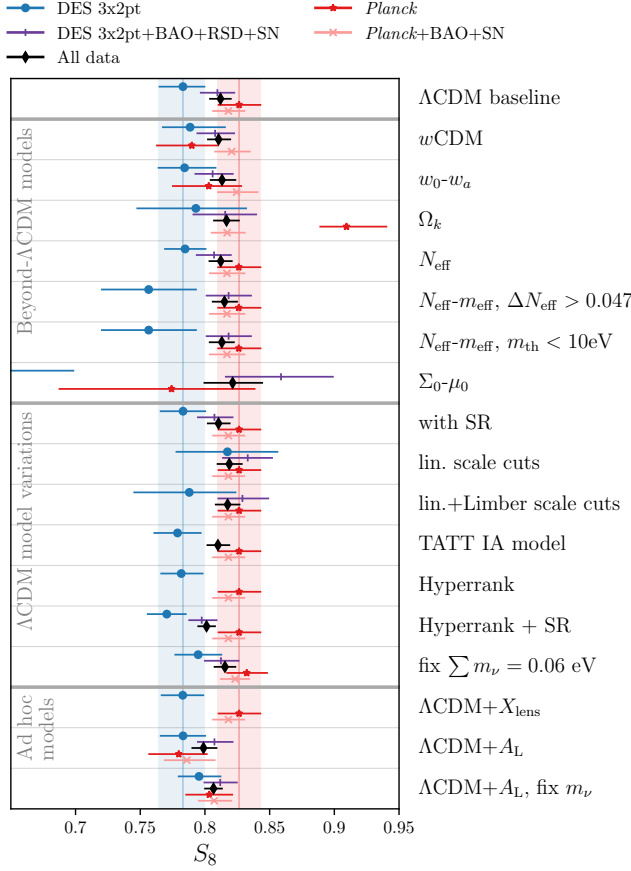


FIG. 13. Constraints on  $S_8$  in different models and under different analysis assumptions. Points and error bars show the means and marginalized 68% confidence, and the shaded bands mark the location of the top row's  $\Lambda$ CDM baseline points for  $3\times 2$ pt and *Planck*. Missing points simply indicate chains that were not run as part of other robustness tests. The rows with ‘SR’ in the label and those varying  $A_L$  include shear ratio as part of the DES  $3\times 2$ pt likelihood, all others do not. The  $\Sigma_0 - \mu_0$   $3\times 2$ pt constraint of  $S_8 = 0.61^{+0.09}_{-0.16}$  is cut off to improve the dynamic range for other points in the plot.

Model	95% upper bound on $\sum m_\nu$ [eV]	
	All External	All data
$\Lambda$ CDM	0.14	0.14
$w$ CDM	0.17	0.19
$w_0-w_a$	0.25	0.26
$\Omega_k$	0.16	0.15
$N_{\text{eff}}$	0.14	0.16
$\Sigma_0-\mu_0$	0.21	0.14
Binned $\sigma_8(z)$	0.30	0.20
$A_L$	0.14	0.19

TABLE V. Impact of the cosmological model on the 95% confidence upper bound on the sum of neutrino masses. The ‘All external’ column reports constraints from the combination of *Planck*+BAO+RSD+SN, while ‘All data’ additionally includes DES  $3\times 2$ pt constraints ( $3\times 2$ pt+SR in the case of binned  $\sigma_8(z)$  and  $A_L$ ).

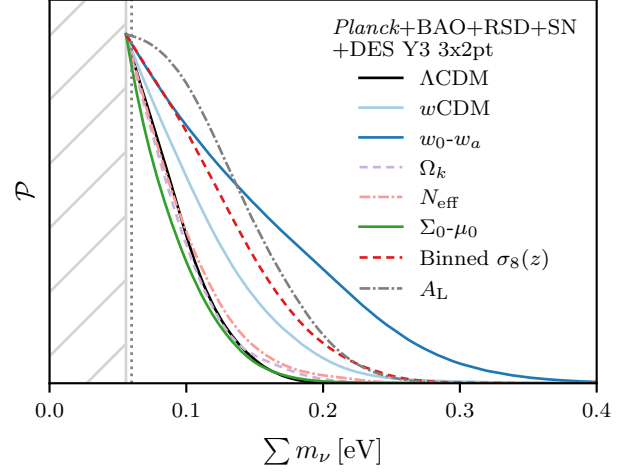


FIG. 14. Marginalized posterior for the sum of neutrino masses obtained from the analysis of DES  $3\times 2$ pt+BAO+RSD+SN+*Planck* in different cosmological models. The dotted vertical gray line shows the minimum mass allowed by neutrino oscillation measurements,  $\sum m_\nu = 0.06$  eV, and the hatched area shows the region excluded by the  $\Omega_\nu h^2 > 0.006$  prior.

*Planck*+BAO+RSD+SN with and without DES  $3\times 2$ pt for several of the models considered in this paper, while Fig. 14 shows the one-dimensional marginalized posterior from the combination of all data. For all models other than  $\Sigma_0 - \mu_0$  and binned  $\sigma_8(z)$ , including the DES  $3\times 2$ pt likelihood either has no effect or slightly weakens the bounds on  $\sum m_\nu$ .

Neutrino mass constraints have a strong dependence on assumptions about the time evolution of dark energy, with the all-data upper bound on  $\sum m_\nu$  increasing relative to the  $\Lambda$ CDM value by a factor of 1.4 for  $w$ CDM and 1.9 for  $w_0 - w_a$ . In contrast,  $\Omega_k$  and  $N_{\text{eff}}$  have little impact. Modifying gravity with  $\Sigma_0 - \mu_0$  weakens external data constraints on neutrino mass by external data, but when DES  $3\times 2$ pt is included the resulting constraint matches that of  $\Lambda$ CDM. Relaxing assumptions about the evolution of structure growth by binning  $\sigma_8(z)$  increases the bound by a factor of 1.4.

## I. Tension between DES and external data for extended models

In Fig. 15 we report measures of tension between DES  $3\times 2$ pt and external low-redshift probes (BAO+RSD+SN), as well as between the combination of all low redshift probes ( $3\times 2$ pt+BAO+RSD+SN) and *Planck*. We evaluate the significance of tension using three statistics, which are discussed in more detail in Appendix E: the Bayes ratio  $R$ , the Suspiciousness  $S$ , and a  $p$  value that converts  $S$  into a tension probability. We define these quantities such that  $\ln R < 0$  and  $\ln S < 0$  correspond to evidence of tension. The quantity  $p(S, d_{\text{BMD}})$ , where BMD stands for ‘Bayesian model dimensionality’ (see Eq. (E6) for more details), approximates the probability, assuming a null hypothesis of agreement between datasets, that

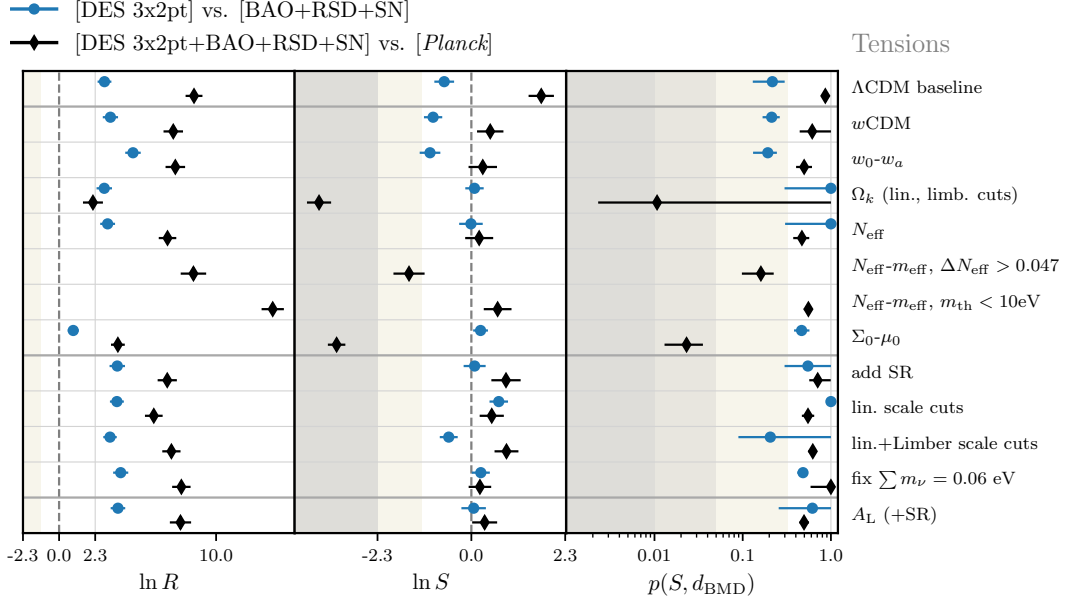


FIG. 15. Measurements of tensions between datasets assuming various cosmological models and analysis variations. Blue circular points denote tensions between DES  $3 \times 2$ pt and low- $z$  BAO, RSD, and SN data, while black diamonds show tensions the combination of all of that low- $z$  data ( $3 \times 2$ pt+BAO+RSD+SN) to Planck CMB constraints. For all panels, points further to the left indicate greater tension. The statistics  $\ln R$  and  $\ln S$  are the Bayesian evidence ratio and Suspiciousness, respectively. Their points and uncertainties reflect the mean and standard deviation of realizations estimating sample variance using the ANESTHETIC software, and shaded regions correspond to substantial and strong evidence of tension according to the Jeffreys' scale. (The Jeffrey's scale is relevant for Suspiciousness because  $\ln S$  can be interpreted as the value of  $\ln R$  associated with the narrowest choice of uninformative prior.) The quantity  $p$  measures the significance accounting for the number of constrained parameters via the Bayesian model dimensionality  $d_{\text{BMD}}$ . In the plot,  $p$ -value errors indicate the 0.16 and 0.84 quantiles of the sample variance realizations, points indicate the mean, and the shaded regions highlight probabilities associated with  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  tension. If error bars are not seen they are smaller than the size of the marker. Further information and definitions of these tension metrics can be found in Appendix E.

we will find a value of  $\ln S$  as low or lower than the observed value. Thus, small  $p$  corresponds to stronger tension. This probability is assessed using both  $S$  and the quantity  $d_{\text{BMD}}$ , which is a Bayesian estimate of the number of directions in parameter space in which a tension could be meaningfully detected — that is, which are constrained by both datasets independently. We show multiple statistics here because while  $R$  is likely to be the most familiar to readers, it has an undesirable sensitivity to the choice of flat prior ranges. In contrast,  $S$  is insensitive to the prior range of well-constrained parameters, and its significance assessed via  $p(S, d_{\text{BMD}})$  is expected to agree with a number of other proposed tension metrics [158].

We estimate  $\ln R$ ,  $\ln S$  and  $d_{\text{BMD}}$  using the ANESTHETIC software<sup>19</sup> [189], which produces an ensemble of 200 realizations capturing the uncertainty introduced by sampling variance. For  $\ln R$  and  $\ln S$ , in Fig. 15 we report the mean of this ensemble, and use error bars (which are occasionally smaller than the datapoint) to show the standard deviation. For the  $p$ -values, whose ensemble distribution is significantly non-Gaussian, we report the median, and approximate its one-

sigma sampling variance errors using the 16 and 84% quantiles. We use the threshold of  $p(S, d_{\text{BMD}}) \geq 0.01$  as a requirement for reporting combined constraints.

For all models and statistics, there is no indication of any tension between DES  $3 \times 2$ pt and the external low-redshift probes (BAO+RSD+SN). This is also true for almost all evaluations of tension between the combination of all low-redshift probes  $3 \times 2$ pt+BAO+RSD+SN and *Planck*. The only cases where we find significant tension are for the  $\Omega_k$  and  $\Sigma_0 - \mu_0$  comparison of  $3 \times 2$ pt+BAO+RSD+SN versus *Planck*, both of which have a significance between 2-3 $\sigma$ .

As was noted above in Sec. VIB, for  $\Omega_k$  the  $p$ -value median is 0.010, exactly at our threshold for reporting combined constraints. This merits further discussion, because in addition to being the most significant measure of tension reported, it is also the noisiest. The 16% and 84% quantiles are 0.002 and 1.0, respectively<sup>20</sup>. This means that at an approximately  $1\sigma$  level of certainty, our evaluation of tension between *Planck* and low-redshift  $\Omega_k$  constraints could plausibly be consistent

<sup>19</sup> <https://github.com/williamjameshandley/anesthetic>

<sup>20</sup> As we will discuss in Appendix E, we assign  $p = 1$  to realizations with  $d_{\text{BMD}} < 0$ , reasoning that there can be no tension measured if there are no shared parameters in the two datasets' independent constraints.

with both a slightly-greater-than- $3\sigma$  tension and with there being no tension at all. This large scatter is driven by the small value of  $d_{\text{BMD}} = 1.5 \pm 1.6$  (reporting the mean and standard deviation from the sample variance estimate). This small  $d_{\text{BMD}}$  means that there is limited overlap in the parameter directions constrained by *Planck* and  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}$ , making the assessment of tension extremely sensitive to noise in the posterior estimates. To further contextualize this finding, we note that the *Planck*-only preference for  $\Omega_k < 0$  driving this tension signal has been the subject of extensive discussion in the literature (see e.g. [11, 160–166]) which highlights the fact that the interpretation of this tension can depend on subtleties related the choice of priors, parameters sampled, *Planck* likelihood calculation method, as well as the relationship to features in the *Planck* power spectra also captured by phenomenological parameter  $A_L$ .

For modified gravity, we see a tension between *Planck* and the other data that is likely driven by the same  $A_L$ -like features of the CMB power spectrum. For  $\Sigma_0 - \mu_0$ , the tension measurement is less significant and much less noisy than for  $\Omega_k$ : the median Suspiciousness  $p$ -value is 0.024 with 16% and 84% quantiles of 0.013 and 0.039.

Note that Fig. 15 does not show tension results for the binned  $\sigma_8(z)$  model. This is because, as described in Sec. III F, in that model we sample different sets of parameters when fitting DES and *Planck* constraints separately and in combination. This makes tension metrics difficult to evaluate, so for simplicity we will show combined ‘all-data’ constraints on binned  $\sigma_8(z)$  without checking a tension metric. This should be a reasonably safe choice because that model’s  $A_i^{P_{\text{lin}}}$  and  $A_{\text{CMB}}^{P_{\text{lin}}}$  parameters introduce enough modeling freedom to capture any differences between observables.

## J. Assessing the preference for extended models relative to $\Lambda\text{CDM}$

Fig. 16 shows several model comparison statistics. We show a variety of metrics here because it allows us to compare the results of different model comparison tests, and to account for the fact that readers may have different preferences regarding which of these tests are most familiar or interpretable. Points further to the left of each subpanel indicate  $\Lambda\text{CDM}$  to be more disfavored with respect to the extended model, while those on the right side of the panels favor  $\Lambda\text{CDM}$  and disfavor the extension. Definitions of these metrics and details about how they are computed can be found in Appendix F, though we will summarize them here. The metrics include the Bayesian evidence ratio  $R$  and Suspiciousness  $S$ , both defined so that  $\ln R < 0$  and  $\ln S < 0$  indicates that the data disfavor  $\Lambda\text{CDM}$ . We also report the ratio of the change in the maximum posterior goodness-of-fit to the number of added parameters,  $\Delta\chi^2/\Delta k$ ,  $\Delta\text{AIC}$  which is an information-theory derived metric based on  $\Delta\chi^2$  with a penalty for adding parameters, and  $\Delta\text{DIC}$  which is related to  $\Delta\text{AIC}$  but adjusted for the number of parameters constrained by the data.

For Suspiciousness, we report two  $p$ -values converting  $\ln S$  to probabilities:  $p(S, d_{\text{BMD}})$ , in which  $d_{\text{BMD}}$  uses Bayesian

model dimensionality to quantify the number of additional parameters constrained in the beyond- $\Lambda\text{CDM}$  analysis, and  $p(S, \Delta k)$  which instead simply uses the number of additional sampled parameters  $\Delta k$  (e.g.  $\Delta k = 1$  for  $\Omega_k$ ,  $\Delta k = 2$  for  $w_0 - w_a$ , etc.). The model comparison definition of  $S$  can be shown to be equivalent to the change in posterior-averaged log-likelihood for different models (see Ref. [154], and for the analogous relation in the case of data tensions, Ref. [19]). This means  $p(S, \Delta k)$  can be viewed as a Bayesian analog to evaluating the probability of the best-fit  $\Delta\chi^2/\Delta k$ .

Like  $\Delta\chi^2/\Delta k$ , the information provided by  $\ln S$  for model comparison is inherently asymmetric: the ‘hardening’ of  $\ln S$  against prior choice means that while it can be used to quantify the significance of preference for an extended model, it will never definitively favor the model with fewer parameters. This is in contrast to  $\ln R$ , for which tight constraints around  $\Lambda\text{CDM}$  parameter values relative to the prior range will cause the extended model to be definitively disfavored.

Examining  $R$ , we find that none of the extended models are significantly preferred against  $\Lambda\text{CDM}$ , with the combined data usually mildly or definitively favoring  $\Lambda\text{CDM}$ . The only exception to this is for  $N_{\text{eff}} - m_{\text{eff}}$  when we impose an upper bound on the sterile neutrino’s thermal mass, where according to the Jeffreys scale we find a substantial (but not strong) preference for the extended model. The preference for  $\Lambda\text{CDM}$  reaches a particularly decisive level in the binned  $\sigma_8(z)$  case, unsurprisingly as it is a phenomenological model adding several parameters without much enhancing the overall fit to data. Considering model variations while assuming  $\Lambda\text{CDM}$ , we find that there is neither a preference for varying neutrino mass nor for varying the additional TATT IA parameters. Relative to the  $N_{\text{eff}} - m_{\text{eff}}$  results, the preference against varying active neutrino mass is comparable to that against  $N_{\text{eff}} - m_{\text{eff}}$  when the  $\Delta N_{\text{eff}} > 0.047$  prior is used, but that the sterile neutrino model is more favored when we use the  $m_{\text{th}} < 10$  eV prior. For the ad hoc models, fixing  $X_{\text{Lens}}$  is favored over varying it, in line with previous studies of the MagLim lens sample, and that there is strong but not significant preference for varying  $A_L$ , in line with previous *Planck* analyses [11, 182, 191].

Turning our attention to  $\ln S$ , we see that panel generally reports values lower than  $\ln R$ , reflecting the expectation that Suspiciousness is not able to significantly favor  $\Lambda\text{CDM}$  over extended models. When assessing the significance of  $S$  measurements using the associated  $p$ -values, we find qualitative agreement with the evidence ratio findings. For both the main set of beyond- $\Lambda\text{CDM}$  models and model variations within  $\Lambda\text{CDM}$ , most preferences for the extended models over  $\Lambda\text{CDM}$  are less than  $1\sigma$  significance, and all are less than  $2\sigma$ . The strong  $\ln R$  evidence for  $N_{\text{eff}} - m_{\text{eff}}$  with  $m_{\text{th}} < 10$  eV translates via  $p(S, \Delta k)$  to a preference of slightly over  $1\sigma$  significance.<sup>21</sup> The largest shift between  $R$  and  $S$  values occurs the binned  $\sigma_8(z)$  model, particularly for the baseline model’s

<sup>21</sup> For that model,  $p(S, d_{\text{BMD}})$  is not reported because the prior on  $m_{\text{th}}$  restricts the parameter space in a way that causes the Bayesian model dimensionality of the  $N_{\text{eff}} - m_{\text{eff}}$  posterior to become smaller than that of  $\Lambda\text{CDM}$ , causing the  $p$ -value to become undefined.

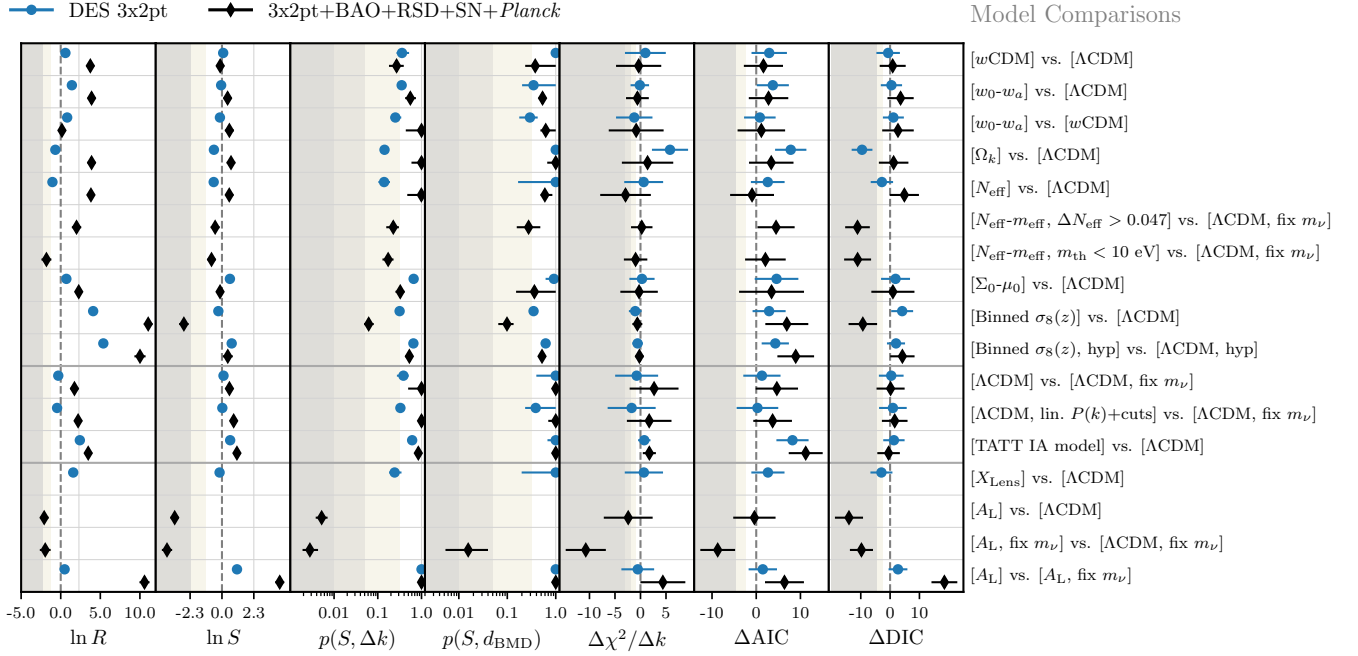


FIG. 16. Model comparison metrics evaluated between pairs of nested models as listed on the right-hand side. All pairs are arranged so that the model listed first has more parameters. In all panels, points further to the left indicate more of a preference for the extended model. Blue points report metrics based on DES Y3-3 $\times$ 2pt constraints alone, while black points are for DES Y3-3 $\times$ 2pt combined with the *Planck*, BAO, RSD, and SN likelihoods. For the Bayes ratio  $\ln R$  and Suspiciousness  $\ln S$ , error bars report the standard deviation associated with sampling variance and shaded regions show regions of substantial and strong preference for the extended parameter space according to the Jeffreys' scale. The two sets of  $p$  values evaluate the significance of the  $\ln S$  results assuming a change in degree of freedom associated with the additional number of sampled parameters  $\Delta k$  in the extended model and the Bayesian model dimensionality  $d_{\text{BMD}}$ . The  $p$ -value points correspond to the mean estimate for sampling variance realizations, the uncertainties correspond to the 0.16 and 0.84 quantiles, and the shaded regions denote probabilities corresponding to  $1\sigma$ ,  $2\sigma$ , and  $\geq 3\sigma$ . For the change in maximum posterior goodness-of-fit,  $\Delta\chi^2$ , errors reflect the propagated standard deviations of 15 MAP estimates performed for each chain, and shaded regions have boundaries at  $\Delta\chi^2/\Delta k = -1, -2$ , and  $-3$ . For the information criterion statistics  $\Delta\text{AIC}$  and  $\Delta\text{DIC}$ , uncertainties reflect the same MAP estimate scatter, and the shaded regions are where model likelihoods quantified by Akaike weights (see e.g. Ref. [190]) match the probabilities associated with the  $\ln R$  Jeffreys scale boundaries. Definitions and more information about these model comparison statistics can be found in Appendix F.

all-data result. However, given the large number of added parameters, the  $p$ -values report that preference for binned  $\sigma_8(z)$  remains insignificant, at less than  $2\sigma$ .

The  $\Delta\text{AIC}$  roughly tracks the evidence ratio, and the  $\Delta\chi^2$  and  $\Delta\text{DIC}$  the Suspiciousness, as expected from the respective definitions.

The overall model comparison conclusion is that none of the models considered offers a compelling alternative to  $\Lambda\text{CDM}$  in explaining the data.

## VII. CONCLUSIONS

We have presented constraints on extensions to the  $\Lambda\text{CDM}$  cosmological model from DES Y3 measurements of cosmic shear, galaxy-galaxy lensing, and clustering (3 $\times$ 2pt summary statistics) in addition to state-of-the-art external data. We investigated how such extensions affect the modeling of 3 $\times$ 2pt observables, and validated the analysis using simulated and blinded real data to ensure that known sources systematic error cannot lead to a false detection of beyond- $\Lambda\text{CDM}$  cosmology.

This work allows us to obtain robust and precise constraints on beyond- $\Lambda\text{CDM}$  cosmology thanks to the unprecedented statistical power of the DES Y3 galaxy METACALIBRATION shape and MagLim lens catalogs. Our analysis indicates no significant deviations from  $\Lambda\text{CDM}$  and its precision is primarily limited by the need for further theoretical developments.

We first expand the exploration of dark energy properties by constraining time dependence of its equation of state. While constraints from DES 3 $\times$ 2pt alone do not contribute significantly to  $w_0 - w_a$  information compared combination of all external data, their precision are comparable to those from other individual cosmological probes. The precision of constraints from DES Y3 3 $\times$ 2pt on  $w_0 - w_a$  is comparable to that of *Planck* alone, and our  $w_a$  constraints are slightly tighter than the latest measurements from Pantheon+ alone [159]. Combining datasets yields precise estimates for  $w_0 - w_a$  which are consistent with a cosmological constant, with the constraining power from measurements of only low redshift 3 $\times$ 2pt +BAO+RSD+SN probes comparable to data combinations including CMB observables.

DES  $3 \times 2$ pt measurements contribute to constraints on the curvature density of the Universe  $\Omega_k$  mainly by constraining  $\Omega_m$ , which helps break a degeneracy between  $\Omega_m$  and  $\Omega_k$  when  $3 \times 2$ pt data are combined with BAO+RSD+SN, leading to a 20% improvement on curvature constraints that can be obtained from low redshift probes. While this low-redshift measurement of  $\Omega_k$  is an order of magnitude weaker than constraints including CMB observables, it is an interesting independent check, given the much-discussed  $\sim 3\sigma$  tension between *Planck* and BAO curvature constraints, which we recover. We find combined constraints to be compatible with flatness, whether or not *Planck* likelihoods are included in the analysis. The constraining power contributed by DES measurements to this study is limited by the lack of validated non-linear LSS modeling and non-Limber projection calculations for non-flat geometry.

Next, we constrain two models sensitive to additional relativistic particle species in the early Universe. We find that DES  $3 \times 2$ pt measurements have little impact on inferences about changes to the number of relativistic species parameterized by  $N_{\text{eff}}$ , but that they are a powerful tool for constraining the impact of light relic particles with non-zero mass on the evolution of large scale structure. We explore this by constraining a species of sterile neutrinos with effective mass  $m_{\text{eff}}$  and a temperature set by  $N_{\text{eff}}$ . As in the case when the parameter is varied alone,  $N_{\text{eff}}$  is primarily constrained by CMB observables, while growth information from DES Y3  $3 \times 2$ pt and external RSD data allows us to tightly constrain  $m_{\text{eff}}$ . In doing this, our combined analysis of all data improves upon the best available constraints on  $m_{\text{eff}}$  by a factor of three, finding  $m_{\text{eff}} < 0.20\text{eV}$ . This constraining power is limited by a lack of validated small scale modeling, which requires us to apply conservative linear scale cuts to the  $3 \times 2$ pt measurements. Given this, we stress that modeling advances such as those being developed in e.g. Refs. [102, 107, 108] will be key to enabling more precise constraints, and that, excitingly, more powerful constraints are attainable even with existing data.

We also test gravity on cosmological scales measuring the  $\Sigma_0 - \mu_0$  parameters. The most interesting constraints for this model come from the combination of multiple observables. In particular, the complementary approach of DES  $3 \times 2$ pt and external RSD measurements to measuring large scale structure allows us to break degeneracies between the modified gravity parameters and  $S_8$ . Their combination thus gives tight constraints, particularly on  $\Sigma_0$ , resulting in consistency with general relativity. We find that *Planck* temperature and polarization constraints prefer slightly higher values of  $\Sigma_0$  than either  $3 \times 2$ pt alone or the combination of all low-redshift probes, likely driven by high- $\ell$  feature of the CMB power spectra that drives the offset seen in the  $\Omega_k$  analysis. When all data are analyzed together the resulting constraints remain consistent with GR. As with sterile neutrinos, the constraining power contributed by DES  $3 \times 2$ pt is also limited by our use of linear scale cuts. In fact, because the increased precision  $3 \times 2$ pt measurements causes our procedure for defining linear scale cuts to remove larger fraction of data points in Y3 compared to Y1, constraints placed on  $\Sigma_0$  by  $3 \times 2$ pt alone to actually weaken compared to the similar analysis in

DES-Y1Ext. Looking ahead to DES Y6 and next-generation surveys, this underlines the need for more sophisticated methods of accounting for non-linear modeling uncertainties when performing cosmological tests of gravity.

Finally, we perform a more generic test of  $\Lambda$ CDM's predictions for structure growth via a binned  $\sigma_8(z)$  model, in which we introduce amplitude parameters that allow the normalization of the matter power spectrum to vary independently in four redshift bins defined by the lens galaxy sample. While constraints on the sampled amplitude parameters are not robust to changes in how we account for source galaxy photometric redshift uncertainties, the  $\sigma_8$  values inferred separately for each redshift bin are more robust, especially for the higher redshift bins when  $3 \times 2$ pt constraints are combined with external data. This analysis finds no significant deviation from the prediction of  $\Lambda$ CDM, and highlights the importance of carefully accounting for the impact of photo- $z$  uncertainties when investigated beyond- $\Lambda$ CDM parameterizations which affect the growth of structure.

In summary, we have conducted robust tests of extensions to  $\Lambda$ CDM using the unprecedentedly precise DES Y3  $3 \times 2$ pt measurements in combination with other state-of-the-art cosmological data, while underlining challenges that will need to be addressed for future wide field galaxy surveys to further test the laws and contents of the Universe. We ultimately detect no significant preference for any of the extended models studied in our analysis. Thus,  $\Lambda$ CDM remains the favored model to describe our data.

## ACKNOWLEDGEMENTS

The analysis made use of the software tools SciPy [192], Astropy [193, 194], Numpy [195, 196], Matplotlib [197], CAMB [198, 199], MG-CAMB [128, 200, 201], GETDIST [202], MULTINEST [203–205], POLYCHORD [75, 76], ANESTHETIC [189], COSMOSIS [206], and GNU PARALLEL [207]. Elements of the DES modeling pipeline additionally use COSMOLIKE [208] including COSMOCOV [73], HALOFIT [81, 82], FAST-PT [88, 209], and NICA EA [210].

This work was supported through computational resources and services provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231; by the Sherlock cluster, supported by Stanford University and the Stanford Research Computing Center; by the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) and by the JPL ITSD High Performance Computing group.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmo-

logical Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas

A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

- 
- [1] A. G. Riess *et al.* (Supernova Search Team), *Astron. J.* **116**, 1009 (1998), [arXiv:astro-ph/9805201](#).
  - [2] S. Perlmutter *et al.* (Supernova Cosmology Project), *Astrophys. J.* **517**, 565 (1999), [arXiv:astro-ph/9812133](#).
  - [3] P. Astier *et al.* (SNLS), *Astron. Astrophys.* **447**, 31 (2006), [arXiv:astro-ph/0510447](#).
  - [4] W. M. Wood-Vasey *et al.* (ESSENCE), *Astrophys. J.* **666**, 694 (2007), [arXiv:astro-ph/0701041](#).
  - [5] J. Guy *et al.*, *A&A* **523**, A7 (2010), [arXiv:1010.4743 \[astro-ph.CO\]](#).
  - [6] A. Conley *et al.*, *Astrophys. J. Supp.* **192**, 1 (2011), [arXiv:1104.1443 \[astro-ph.CO\]](#).
  - [7] A. Rest *et al.*, *Astrophys. J.* **795**, 44 (2014), [arXiv:1310.3828 \[astro-ph.CO\]](#).
  - [8] D. M. Scolnic *et al.*, *Astrophys. J.* **859**, 101 (2018), [arXiv:1710.00845 \[astro-ph.CO\]](#).
  - [9] DES Collaboration, T. M. C. Abbott, *et al.* (DES), *Astrophys. J. Lett.* **872**, L30 (2019), [arXiv:1811.02374 \[astro-ph.CO\]](#).
  - [10] G. Hinshaw *et al.* (WMAP), *Astrophys. J. Suppl.* **208**, 19 (2013), [arXiv:1212.5226 \[astro-ph.CO\]](#).
  - [11] N. Aghanim *et al.* (Planck), *Astron. Astrophys.* **641**, A6 (2020), [Erratum: *Astron. Astrophys.* 652, C4 (2021)], [arXiv:1807.06209 \[astro-ph.CO\]](#).
  - [12] S. Aiola *et al.* (ACT), *JCAP* **12**, 047 (2020), [arXiv:2007.07288 \[astro-ph.CO\]](#).
  - [13] C. Blake *et al.*, *Mon. Not. R. Astron. Soc.* **425**, 405 (2012), [arXiv:1204.3674 \[astro-ph.CO\]](#).
  - [14] J. Elvin-Poole, M. Crocce, A. Ross, *et al.* (DES), *Phys. Rev. D* **98**, 042006 (2018), [arXiv:1708.01536 \[astro-ph.CO\]](#).
  - [15] S. Alam *et al.* (eBOSS), *Phys. Rev. D* **103**, 083533 (2021), [arXiv:2007.08991 \[astro-ph.CO\]](#).
  - [16] C. Heymans *et al.*, *Mon. Not. Roy. Astron. Soc.* **432**, 2433 (2013), [arXiv:1303.1808 \[astro-ph.CO\]](#).
  - [17] M. A. Troxel, N. MacCrann, J. Zuntz, *et al.* (DES), *Phys. Rev. D* **98**, 043528 (2018), [arXiv:1708.01538 \[astro-ph.CO\]](#).
  - [18] C. Hikage *et al.* (HSC), *Publ. Astron. Soc. Jap.* **71**, 43 (2019), [arXiv:1809.09148 \[astro-ph.CO\]](#).
  - [19] C. Heymans *et al.*, *Astron. Astrophys.* **646**, A140 (2021), [arXiv:2007.15632 \[astro-ph.CO\]](#).
  - [20] V. Bonvin *et al.*, *Mon. Not. Roy. Astron. Soc.* **465**, 4914 (2017), [arXiv:1607.01790 \[astro-ph.CO\]](#).
  - [21] D. Huterer, D. Shafer, D. Scolnic, and F. Schmidt, *JCAP* **05**, 015 (2017), [arXiv:1611.09862 \[astro-ph.CO\]](#).
  - [22] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, *Phys. Rev. D* **101**, 083504 (2020), [arXiv:1912.08208 \[astro-ph.CO\]](#).
  - [23] J. Frieman, M. Turner, and D. Huterer, *Ann. Rev. Astron. Astrophys.* **46**, 385 (2008), [arXiv:0803.0982 \[astro-ph\]](#).
  - [24] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Roza, *Phys. Rept.* **530**, 87 (2013), [arXiv:1201.2434 \[astro-ph.CO\]](#).
  - [25] S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).
  - [26] S. M. Carroll, *Living Rev. Rel.* **4**, 1 (2001), [arXiv:astro-ph/0004075 \[astro-ph\]](#).
  - [27] J. Martin, *Comptes Rendus Physique* **13**, 566 (2012), [arXiv:1205.3365 \[astro-ph.CO\]](#).
  - [28] A. Padilla, (2015), [arXiv:1502.05296 \[hep-th\]](#).
  - [29] B. Flaugher, H. T. Diehl, K. Honscheid, *et al.* (DES), *Astro. Journal* **150**, 150 (2015), [arXiv:1504.02900 \[astro-ph.IM\]](#).
  - [30] DES Collaboration (DES), *Phys. Rev. D* **98**, 043526 (2018), [arXiv:1708.01530 \[astro-ph.CO\]](#).
  - [31] DES Collaboration (DES), *Mon. Not. Roy. Astron. Soc.* **460**, 1270 (2016), [arXiv:1601.00329 \[astro-ph.CO\]](#).

- [32] H. Aihara *et al.*, Publ. Astron. Soc. Jap. **70**, S4 (2018), arXiv:1704.05858 [astro-ph.IM].
- [33] K. Kuijken *et al.*, Mon. Not. Roy. Astron. Soc. **454**, 3500 (2015), arXiv:1507.00738 [astro-ph.CO].
- [34] K. S. Dawson *et al.*, Astron. J. **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO].
- [35] M. E. Levi *et al.* (DESI), (2019), arXiv:1907.10688 [astro-ph.IM].
- [36] L. Verde, T. Treu, and A. G. Riess, Nature Astron. **3**, 891 (2019), arXiv:1907.10625 [astro-ph.CO].
- [37] L. Knox and M. Millea, Phys. Rev. D **101**, 043533 (2020), arXiv:1908.03663 [astro-ph.CO].
- [38] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, Class. Quant. Grav. **38**, 153001 (2021), arXiv:2103.01183 [astro-ph.CO].
- [39] P. Shah, P. Lemos, and O. Lahav, Astron. Astrophys. Rev. **29**, 9 (2021), arXiv:2109.01161 [astro-ph.CO].
- [40] A. G. Riess *et al.*, Astrophys. J. Lett. **934**, L7 (2022), arXiv:2112.04510 [astro-ph.CO].
- [41] E. Abdalla *et al.*, JHEAp **34**, 49 (2022), arXiv:2203.06142 [astro-ph.CO].
- [42] M. Asgari *et al.* (KiDS), Astron. Astrophys. **645**, A104 (2021), arXiv:2007.15633 [astro-ph.CO].
- [43] A. Krolewski, S. Ferraro, and M. White, JCAP **12**, 028 (2021), arXiv:2105.03421 [astro-ph.CO].
- [44] DES Collaboration (DES), Phys. Rev. D **105**, 023520 (2022), arXiv:2105.13549 [astro-ph.CO].
- [45] C. Chang, Y. Omori, E. J. Baxter, C. Doux, A. Choi, S. Pandey, A. Alarcon, O. Alves, A. Amon, F. Andrade-Oliveira, *et al.* (DES, SPT), (2022), arXiv:2203.12440 [astro-ph.CO].
- [46] DES and SPT Collaborations (DES, SPT), (2022), arXiv:2206.10824 [astro-ph.CO].
- [47] C. Chang *et al.* (LSST Dark Energy Science), Mon. Not. Roy. Astron. Soc. **482**, 3696 (2019), arXiv:1808.07335 [astro-ph.CO].
- [48] A. Leauthaud *et al.* (DES), Mon. Not. Roy. Astron. Soc. **510**, 6150 (2022), arXiv:2111.13805 [astro-ph.CO].
- [49] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey, S. Avila, O. Ballester, M. Banerji, W. Barkhouse, *et al.*, The Astrophysical Journal Supplement Series **239**, 18 (2018).
- [50] DES Collaboration (DES), Phys. Rev. D **99**, 123505 (2019), arXiv:1810.02499 [astro-ph.CO].
- [51] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A5 (2020), arXiv:1907.12875 [astro-ph.CO].
- [52] B. Flaugher *et al.* (DES), The Astronomical Journal **150**, 150 (2015), arXiv:1504.02900 [astro-ph.IM].
- [53] I. Sevilla-Noarbe *et al.* (DES), Astrophys. J. Suppl. **254**, 24 (2021), arXiv:2011.03407 [astro-ph.CO].
- [54] M. Gatti, E. Sheldon, *et al.* (DES), Mon. Not. Roy. Astron. Soc. **504**, 4312 (2021), arXiv:2011.03408 [astro-ph.CO].
- [55] E. Huff and R. Mandelbaum, (2017), arXiv:1702.02600 [astro-ph.CO].
- [56] E. S. Sheldon and E. M. Huff, Astrophys. J. **841**, 24 (2017), arXiv:1702.02601 [astro-ph.CO].
- [57] M. Jarvis *et al.* (DES), Mon. Not. Roy. Astron. Soc. **501**, 1282 (2021), arXiv:2011.03409 [astro-ph.IM].
- [58] N. MacCrann *et al.* (DES), Mon. Not. Roy. Astron. Soc. **509**, 3371 (2021), arXiv:2012.08567 [astro-ph.CO].
- [59] J. T. Myles, A. Alarcon, *et al.* (DES), Mon. Not. Roy. Astron. Soc. **505**, 4249 (2021), arXiv:2012.08566 [astro-ph.CO].
- [60] R. Buchs, C. Davis, *et al.* (DES), Mon. Not. Roy. Astron. Soc. **489**, 820 (2019), arXiv:1901.05005 [astro-ph.CO].
- [61] W. G. Hartley *et al.* (DES), Mon. Not. Roy. Astron. Soc. **509**, 3547 (2021), arXiv:2012.12824 [astro-ph.CO].
- [62] S. Everett *et al.* (DES), Astrophys. J. Suppl. **258**, 15 (2022), arXiv:2012.12825 [astro-ph.CO].
- [63] A. Porredon *et al.* (DES), Phys. Rev. D **103**, 043503 (2021), arXiv:2011.03411 [astro-ph.CO].
- [64] R. Cawthon *et al.* (DES), Mon. Not. Roy. Astron. Soc. **513**, 5517 (2022), arXiv:2012.12826 [astro-ph.CO].
- [65] M. Rodríguez-Monroy *et al.* (DES), Mon. Not. Roy. Astron. Soc. **511**, 2665 (2022), arXiv:2105.13540 [astro-ph.CO].
- [66] A. Porredon *et al.* (DES), (2021), arXiv:2105.13546 [astro-ph.CO].
- [67] A. Amon *et al.* (DES), Phys. Rev. D **105**, 023514 (2022), arXiv:2105.13543 [astro-ph.CO].
- [68] L. Secco, S. Samuroff, *et al.* (DES), Phys. Rev. D **105**, 023515 (2022), arXiv:2105.13544 [astro-ph.CO].
- [69] J. Prat *et al.* (DES), Phys. Rev. D **105**, 083528 (2022), arXiv:2105.13541 [astro-ph.CO].
- [70] M. Jarvis, G. Bernstein, and B. Jain, Mon. Not. Roy. Astron. Soc. **352**, 338 (2004), arXiv:astro-ph/0307393 [astro-ph].
- [71] E. Krause *et al.* (DES), (2021), arXiv:2105.13548 [astro-ph.CO].
- [72] C. Sánchez, J. Prat, *et al.* (DES), Phys. Rev. D **105**, 083529 (2022), arXiv:2105.13542 [astro-ph.CO].
- [73] X. Fang, T. Eifler, and E. Krause, Mon. Not. Roy. Astron. Soc. **497**, 2699 (2020), arXiv:2004.04833 [astro-ph.CO].
- [74] O. Friedrich *et al.* (DES), Mon. Not. R. Astron. Soc. **508**, 3125 (2021), arXiv:2012.08568 [astro-ph.CO].
- [75] W. J. Handley, M. P. Hobson, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **450**, L61 (2015), arXiv:1502.01856.
- [76] W. J. Handley, M. P. Hobson, and A. N. Lasenby, Mon. Not. R. Astron. Soc. **453**, 4384 (2015), arXiv:1506.00171 [astro-ph.IM].
- [77] P. Lemos, N. Weaverdyck, *et al.* (DES), Submitted to Mon. Not. R. Astron. Soc. (2022), 10.1093/mnras/stac2786, arXiv:2202.08233 [astro-ph.CO].
- [78] A. Stebbins, (1996), arXiv:astro-ph/9609149.
- [79] D. N. Limber, Astrophys. J. **117**, 134 (1953).
- [80] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchmann (VIRGO Consortium), Mon. Not. Roy. Astron. Soc. **341**, 1311 (2003), arXiv:astro-ph/0207664 [astro-ph].
- [81] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Astrophys. J. **761**, 152 (2012), arXiv:1208.2701 [astro-ph.CO].
- [82] S. Bird, M. Viel, and M. G. Haehnelt, Mon. Not. Roy. Astron. Soc. **420**, 2551 (2012), arXiv:1109.4416 [astro-ph.CO].
- [83] X. Fang, E. Krause, T. Eifler, and N. MacCrann, JCAP **05**, 010 (2020), arXiv:1911.11947 [astro-ph.CO].
- [84] E. Krause, E. Eifler, J. Zuntz, O. Friedrich, M. Troxel, *et al.* (DES), Submitted to: Phys. Rev. D (2017), arXiv:1706.09359 [astro-ph.CO].
- [85] C. M. Hirata and U. c. v. Seljak, Phys. Rev. D **70**, 063526 (2004).
- [86] S. Bridle and L. King, New Journal of Physics **9**, 444 (2007).
- [87] J. Blazek, Z. Vlah, and U. Seljak, Journal of Cosmology and Astroparticle Physics **2015**, 015 (2015).
- [88] J. E. McEwen, X. Fang, C. M. Hirata, and J. A. Blazek, JCAP **09**, 015 (2016), arXiv:1603.04826 [astro-ph.CO].
- [89] J. Elvin-Poole *et al.* (DES), (2022), arXiv:2209.09782 [astro-ph.CO].
- [90] B. Jain and A. Taylor, Phys. Rev. Lett. **91**, 141302 (2003), arXiv:astro-ph/0306046.

- [91] J. Prat, C. Sanchez, Y. Fang, *et al.* (DES), *Phys. Rev. D* **98**, 042005 (2018), arXiv:1708.01537 [astro-ph.CO].
- [92] R. Terasawa, R. Takahashi, T. Nishimichi, and M. Takada, *Phys. Rev. D* **106**, 083504 (2022).
- [93] P. Ade *et al.* (Planck), *Astron. Astrophys.* **594**, A14 (2016), arXiv:1502.01590 [astro-ph.CO].
- [94] O. Pisanti, A. Cirillo, S. Esposito, F. Iocco, G. Mangano, G. Miele, and P. D. Serpico, *Comput. Phys. Commun.* **178**, 956 (2008), arXiv:0705.0290 [astro-ph].
- [95] E. V. Linder, *Phys. Rev. Lett.* **90**, 091301 (2003), astro-ph/0208512.
- [96] M. Chevallier and D. Polarski, *Int. J. Mod. Phys. D* **10**, 213 (2001), arXiv:gr-qc/0009008 [gr-qc].
- [97] D. Huterer and M. S. Turner, *Phys. Rev. D* **64**, 123527 (2001), arXiv:astro-ph/0012510.
- [98] J. J. Bennett, G. Buldgen, P. F. De Salas, M. Drewes, S. Gariazzo, S. Pastor, and Y. Y. Y. Wong, *JCAP* **04**, 073 (2021), arXiv:2012.02726 [hep-ph].
- [99] K. Akita and M. Yamaguchi, *JCAP* **08**, 012 (2020), arXiv:2005.07047 [hep-ph].
- [100] A. Lewis, “CAMB Notes,” (2014).
- [101] W. L. Xu, J. B. Muñoz, and C. Dvorkin, *Phys. Rev. D* **105**, 095029 (2022), arXiv:2107.09664 [astro-ph.CO].
- [102] J. B. Muñoz and C. Dvorkin, *Phys. Rev. D* **98**, 043503 (2018), arXiv:1805.11623 [astro-ph.CO].
- [103] S. Colombi, S. Dodelson, and L. M. Widrow, *Astrophys. J.* **458**, 1 (1996), arXiv:astro-ph/9505029.
- [104] J. Lesgourgues, G. Mangano, G. Miele, and S. Pastor, *Neutrino Cosmology* (Cambridge University Press, 2013).
- [105] M. LoVerde, *Phys. Rev. D* **90**, 083530 (2014), arXiv:1405.4855 [astro-ph.CO].
- [106] A. Aviles, A. Banerjee, G. Niz, and Z. Slepian, *JCAP* **11**, 028 (2021).
- [107] J. Brandbyge and S. Hannestad, *JCAP* **10**, 015 (2017), arXiv:1706.00025 [astro-ph.CO].
- [108] A. Banerjee, S. Das, A. Maharana, and R. K. Sharma, *Mon. Not. Roy. Astron. Soc.* **516**, 2038 (2022), arXiv:2202.09840 [astro-ph.CO].
- [109] R. Caldwell, A. Cooray, and A. Melchiorri, *Physical Review D* **76** (2007), 10.1103/physrevd.76.023507.
- [110] L. Amendola, M. Kunz, and D. Sapone, *Journal of Cosmology and Astroparticle Physics* **2008**, 013 (2008), arXiv:0704.2421.
- [111] W. Hu and I. Sawicki, *Physical Review D* **76**, 104043 (2007), publisher: American Physical Society.
- [112] B. Jain and P. Zhang, *Phys. Rev. D* **78**, 063503 (2008), arXiv:0709.2375 [astro-ph].
- [113] S. F. Daniel, R. R. Caldwell, A. Cooray, and A. Melchiorri, *Physical Review D* **77**, 103513 (2008), publisher: American Physical Society.
- [114] E. Bertschinger and P. Zukin, *Physical Review D* **78**, 024015 (2008), arXiv:0801.2431.
- [115] G.-B. Zhao, L. Pogosian, A. Silvestri, and J. Zylberberg, *Phys. Rev. D* **79**, 083513 (2009).
- [116] L. Pogosian, A. Silvestri, K. Koyama, and G.-B. Zhao, *Physical Review D* **81**, 104023 (2010), publisher: American Physical Society.
- [117] G.-B. Zhao, T. Giannantonio, L. Pogosian, A. Silvestri, D. J. Bacon, K. Koyama, R. C. Nichol, and Y.-S. Song, *Phys. Rev. D* **81**, 103510 (2010), arXiv:1003.0001 [astro-ph.CO].
- [118] A. Silvestri, L. Pogosian, and R. V. Buniy, *Phys. Rev. D* **87**, 104015 (2013), arXiv:1302.1193 [astro-ph.CO].
- [119] F. Simpson *et al.*, *Mon. Not. Roy. Astron. Soc.* **429**, 2249 (2013), arXiv:1212.3339 [astro-ph.CO].
- [120] A. Ferté, D. Kirk, A. R. Liddle, and J. Zuntz, *Phys. Rev. D* **99**, 083512 (2019), arXiv:1712.01846 [astro-ph.CO].
- [121] S. Joudaki, A. Mead, C. Blake, A. Choi, J. de Jong, T. Erben, I. Fenech Conti, R. Herbonnet, C. Heymans, H. Hildebrandt, H. Hoekstra, B. Joachimi, D. Klaes, F. Köhlinger, K. Kuijken, J. McFarland, L. Miller, P. Schneider, and M. Viola, *Monthly Notices of the Royal Astronomical Society* **471**, 1259 (2017), <https://academic.oup.com/mnras/article-pdf/471/2/1259/19400712/stx998.pdf>.
- [122] C. Garcia-Quintero, M. Ishak, and O. Ning, *JCAP* **12**, 018 (2020), arXiv:2010.12519 [astro-ph.CO].
- [123] A. Ferté *et al.* (DES), (in prep).
- [124] D. B. Thomas, *Physical Review D* **101** (2020), 10.1103/physrevd.101.123517.
- [125] M. Cataneo, L. Lombriser, C. Heymans, A. Mead, A. Barreira, S. Bose, and B. Li, *Mon. Not. Roy. Astron. Soc.* **488**, 2121 (2019), arXiv:1812.05594 [astro-ph.CO].
- [126] B. Bose, M. Cataneo, T. Tröster, Q. Xia, C. Heymans, and L. Lombriser, *Mon. Not. Roy. Astron. Soc.* **498**, 4650 (2020), arXiv:2005.12184 [astro-ph.CO].
- [127] A. Chen *et al.* (DES), *Phys. Rev. D* **103**, 123528 (2021), arXiv:2011.04606 [astro-ph.CO].
- [128] A. Zucca, L. Pogosian, A. Silvestri, and G. Zhao, *JCAP* **05**, 001 (2019), arXiv:1901.05956 [astro-ph.CO].
- [129] J. A. Blazek, N. MacCrann, M. A. Troxel, and X. Fang, *Phys. Rev. D* **100**, 103506 (2019), arXiv:1708.09247 [astro-ph.CO].
- [130] J. Muir *et al.* (DES), *Phys. Rev. D* **103**, 023528 (2021), arXiv:2010.05924 [astro-ph.CO].
- [131] E. Calabrese, A. Slosar, A. Melchiorri, G. F. Smoot, and O. Zahn, *Phys. Rev. D* **77**, 123531 (2008), arXiv:0803.2309 [astro-ph].
- [132] C. Garcia-Quintero and M. Ishak, *Mon. Not. R. Astron. Soc.* **506**, 1704 (2021), arXiv:2009.01189 [astro-ph.CO].
- [133] E. V. Linder, *JCAP* **10**, 042 (2020), arXiv:2003.10453 [astro-ph.CO].
- [134] H. Prince and J. Dunkley, *Physical Review D* **100** (2019), 10.1103/physrevd.100.083502.
- [135] R. Ahumada *et al.* (SDSS-IV), *Astrophys. J. Suppl.* **249**, 3 (2020), arXiv:1912.02905 [astro-ph.GA].
- [136] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, *Mon. Not. Roy. Astron. Soc.* **416**, 3017 (2011), arXiv:1106.3366 [astro-ph.CO].
- [137] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, *Mon. Not. Roy. Astron. Soc.* **449**, 835 (2015), arXiv:1409.3242 [astro-ph.CO].
- [138] A. Tamone, A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, E. Burtin, V. Ruhlmann-Kleider, A. J. Ross, S. Alam, W. J. Percival, and *et al.*, *Monthly Notices of the Royal Astronomical Society* **499**, 5527–5546 (2020).
- [139] A. Barreira, A. G. Sánchez, and F. Schmidt, *Phys. Rev. D* **94**, 084022 (2016), arXiv:1605.03965 [astro-ph.CO].
- [140] J. Muir *et al.* (DES), *Mon. Not. Roy. Astron. Soc.* **494**, 4454 (2020), arXiv:1911.05929 [astro-ph.CO].
- [141] E. Rozo *et al.* (DES), *Mon. Not. Roy. Astron. Soc.* **461**, 1431 (2016), arXiv:1507.05460 [astro-ph.IM].
- [142] S. Pandey *et al.* (DES Collaboration), *Phys. Rev. D* **106**, 043520 (2022), arXiv:2105.13545 [astro-ph.CO].
- [143] G. Efstathiou and J. R. Bond, *Mon. Not. Roy. Astron. Soc.* **304**, 75 (1999), arXiv:astro-ph/9807103.
- [144] H.-J. Huang *et al.* (DES), *Mon. Not. Roy. Astron. Soc.* **502**, 6010 (2021), arXiv:2007.15026 [astro-ph.CO].
- [145] H.-J. Huang, T. Eifler, R. Mandelbaum, and S. Do-

- delson, *Mon. Not. Roy. Astron. Soc.* **488**, 1652 (2019), arXiv:1809.01146 [astro-ph.CO].
- [146] J. Schaye, C. Dalla Vecchia, C. M. Booth, R. P. C. Wiersma, T. Theuns, M. R. Haas, S. Bertone, A. R. Duffy, I. G. McCarthy, and F. van de Voort, *Mon. Not. Roy. Astron. Soc.* **402**, 1536 (2010), arXiv:0909.5196 [astro-ph.CO].
- [147] M. P. van Daalen, J. Schaye, C. M. Booth, and C. Dalla Vecchia, *Monthly Notices of the Royal Astronomical Society* **415**, 3649–3665 (2011).
- [148] C. M. Booth and J. Schaye, *Monthly Notices of the Royal Astronomical Society* **398**, 53–74 (2009).
- [149] P. McDonald and A. Roy, *Journal of Cosmology and Astroparticle Physics* **2009**, 020–020 (2009).
- [150] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, *Physical Review D* **86** (2012), 10.1103/physrevd.86.083540.
- [151] K. C. Chan, R. Scoccimarro, and R. K. Sheth, *Phys. Rev. D* **85**, 083509 (2012), arXiv:1201.3614 [astro-ph.CO].
- [152] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, and P. McDonald, *Phys. Rev. D* **90**, 123522 (2014), arXiv:1405.1447 [astro-ph.CO].
- [153] M. Knabenhans et al. (Euclid), *Mon. Not. Roy. Astron. Soc.* **484**, 5509 (2019), arXiv:1809.04695 [astro-ph.CO].
- [154] N. Weaverdyck, O. Alves, et al., (in prep.).
- [155] J. P. Cordero et al. (DES), *Mon. Not. R. Astron. Soc.* **511**, 2170 (2022), arXiv:2109.09636 [astro-ph.CO].
- [156] C. Sánchez, M. Raveri, A. Alarcon, and G. M. Bernstein, *Mon. Not. Roy. Astron. Soc.* **498**, 2984 (2020), arXiv:2004.09542 [astro-ph.CO].
- [157] C. Sánchez and G. M. Bernstein, *Mon. Not. Roy. Astron. Soc.* **483**, 2801 (2019), arXiv:1807.11873 [astro-ph.CO].
- [158] P. Lemos et al. (DES), *Mon. Not. Roy. Astron. Soc.* **505**, 6179 (2021), arXiv:2012.09554 [astro-ph.CO].
- [159] D. Brout et al., *Astrophys. J.* **938**, 110 (2022), arXiv:2202.04077 [astro-ph.CO].
- [160] C.-G. Park and B. Ratra, *Astrophys. J.* **882**, 158 (2019), arXiv:1801.00213 [astro-ph.CO].
- [161] W. Handley, *Phys. Rev. D* **103**, L041301 (2021), arXiv:1908.09139 [astro-ph.CO].
- [162] E. Di Valentino, A. Melchiorri, and J. Silk, *Nature Astron.* **4**, 196 (2019), arXiv:1911.02087 [astro-ph.CO].
- [163] E. Di Valentino, A. Melchiorri, and J. Silk, *Astrophys. J. Lett.* **908**, L9 (2021), arXiv:2003.04935 [astro-ph.CO].
- [164] G. Efstathiou and S. Gratton, *Mon. Not. Roy. Astron. Soc.* **496**, L91 (2020), arXiv:2002.06892 [astro-ph.CO].
- [165] A. Glanville, C. Howlett, and T. M. Davis, (2022), arXiv:2205.05892 [astro-ph.CO].
- [166] L. T. Hergt, F. J. Agocs, W. J. Handley, M. P. Hobson, and A. N. Lasenby, *Finite inflation in curved space*, Other thesis (2022), arXiv:2205.07374 [astro-ph.CO].
- [167] S. Alam et al. (eBOSS), *Phys. Rev. D* **103**, 083533 (2021), arXiv:2007.08991 [astro-ph.CO].
- [168] J. Bel, J. Larena, R. Maartens, C. Marinoni, and L. Perenon, *JCAP* **09**, 076 (2022), arXiv:2206.03059 [astro-ph.CO].
- [169] R. Sachs and A. Wolfe, *Astrophys. J.* **147**, 73 (1967).
- [170] W. Hu and N. Sugiyama, *Phys. Rev. D* **50**, 627 (1994), arXiv:astro-ph/9310046.
- [171] M. White et al., *JCAP* **02**, 007 (2022), arXiv:2111.09898 [astro-ph.CO].
- [172] S. Brieden, H. Gil-Marín, and L. Verde, *JCAP* **08**, 024 (2022), arXiv:2204.11868 [astro-ph.CO].
- [173] C. García-García, J. R. Zapatero, D. Alonso, E. Bellini, P. G. Ferreira, E.-M. Mueller, A. Nicola, and P. Ruiz-Lapuente, *JCAP* **10**, 030 (2021), arXiv:2105.12108 [astro-ph.CO].
- [174] G. Franco Abellán, R. Murgia, V. Poulin, and J. Lavalley, *Phys. Rev. D* **105**, 063525 (2022), arXiv:2008.09615 [astro-ph.CO].
- [175] Z. Davari, V. Marra, and M. Malekjani, *Mon. Not. Roy. Astron. Soc.* **491**, 1920 (2020), arXiv:1911.00209 [gr-qc].
- [176] J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez, and C. Moreno-Pulido, *Class. Quant. Grav.* **37**, 245003 (2020), arXiv:2006.04273 [astro-ph.CO].
- [177] J. Solà Peracaula, A. Gomez-Valent, J. de Cruz Pérez, and C. Moreno-Pulido, *Astrophys. J. Lett.* **886**, L6 (2019), arXiv:1909.02554 [astro-ph.CO].
- [178] S. Camera, M. Martinelli, and D. Bertacca, *Phys. Dark Univ.* **23**, 100247 (2019), arXiv:1704.06277 [astro-ph.CO].
- [179] E. Di Valentino and S. Bridle, *Symmetry* **10**, 585 (2018).
- [180] R. Murgia, G. F. Abellán, and V. Poulin, *Phys. Rev. D* **103**, 063502 (2021), arXiv:2009.10733 [astro-ph.CO].
- [181] M. Archidiacono, D. C. Hooper, R. Murgia, S. Bohr, J. Lesgourgues, and M. Viel, *JCAP* **10**, 055 (2019), arXiv:1907.01496 [astro-ph.CO].
- [182] E. Di Valentino, A. Melchiorri, and J. Silk, *JCAP* **01**, 013 (2020), arXiv:1908.01391 [astro-ph.CO].
- [183] A. Chudaykin, D. Gorbunov, and N. Nedelko, (2022), arXiv:2203.03666 [astro-ph.CO].
- [184] O. Akarsu, S. Kumar, E. Özüiker, and J. A. Vazquez, *Phys. Rev. D* **104**, 123512 (2021), arXiv:2108.09239 [astro-ph.CO].
- [185] M. Lucca, *Phys. Dark Univ.* **34**, 100899 (2021), arXiv:2105.09249 [astro-ph.CO].
- [186] J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Perez, and C. Moreno-Pulido, *EPL* **134**, 19001 (2021), arXiv:2102.12758 [astro-ph.CO].
- [187] G. Benevento, J. A. Kable, G. E. Addison, and C. L. Bennett, *Astrophys. J.* **935**, 156 (2022), arXiv:2202.09356 [astro-ph.CO].
- [188] S. Heimersheim, N. Schöneberg, D. C. Hooper, and J. Lesgourgues, *JCAP* **12**, 016 (2020), arXiv:2008.08486 [astro-ph.CO].
- [189] W. Handley, *J. Open Source Softw.* **4**, 1414 (2019), arXiv:1905.04768 [astro-ph.IM].
- [190] A. R. Liddle, *Mon. Not. Roy. Astron. Soc.* **377**, L74 (2007), arXiv:astro-ph/0701113.
- [191] S. Alam et al. (BOSS), *Mon. Not. Roy. Astron. Soc.* **470**, 2617 (2017), arXiv:1607.03155 [astro-ph.CO].
- [192] E. Jones, T. Oliphant, P. Peterson, et al., “ey: Open source scientific tools for Python,” (2001–), [Online; accessed 2022].
- [193] T. P. Robitaille et al. (Astropy), *Astron. Astrophys.* **558**, A33 (2013), arXiv:1307.6212 [astro-ph.IM].
- [194] A. M. Price-Whelan et al., *Astron. J.* **156**, 123 (2018), arXiv:1801.02634.
- [195] C. R. Harris et al., *Nature* **585**, 357 (2020), arXiv:2006.10256 [cs.MS].
- [196] T. E. Oliphant, *A guide to NumPy* (Trelgol Publishing, USA, 2006).
- [197] J. D. Hunter, *Computing in Science & Engineering* **9**, 90 (2007).
- [198] A. Lewis, A. Challinor, and A. Lasenby, *Astrophys. J.* **538**, 473 (2000), arXiv:astro-ph/9911177 [astro-ph].
- [199] C. Howlett, A. Lewis, A. Hall, and A. Challinor, *JCAP* **1204**, 027 (2012), arXiv:1201.3654 [astro-ph.CO].
- [200] A. Hojjati, L. Pogosian, and G.-B. Zhao, *JCAP* **1108**, 005 (2011), arXiv:1106.4543 [astro-ph.CO].
- [201] G.-B. Zhao, L. Pogosian, A. Silvestri, and J. Zylberberg, *Phys. Rev. D* **79**, 083513 (2009), arXiv:0809.3791 [astro-ph].
- [202] A. Lewis, *GetDist* (2019), [Online; accessed 2022], arXiv:1910.13970 [astro-ph.IM].
- [203] F. Feroz and M. P. Hobson, *Mon. Not. Roy. Astron. Soc.* **384**, 449 (2008), arXiv:0704.3704 [astro-ph].

- [204] F. Feroz, M. P. Hobson, and M. Bridges, *Mon. Not. Roy. Astron. Soc.* **398**, 1601 (2009), arXiv:0809.3437 [astro-ph].
- [205] F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt, *Open J. Astrophys.* **2**, 10 (2019), arXiv:1306.2144 [astro-ph.IM].
- [206] J. Zuntz, M. Paterno, E. Jennings, D. Rudd, A. Manzotti, S. Dodelson, S. Bridle, S. Sehrish, and J. Kowalkowski, *Astron. Comput.* **12**, 45 (2015), arXiv:1409.3409 [astro-ph.CO].
- [207] O. Tange, *GNU Parallel 2018* (Ole Tange, 2018).
- [208] E. Krause and T. Eifler, *Mon. Not. Roy. Astron. Soc.* **470**, 2100 (2017), arXiv:1601.05779 [astro-ph.CO].
- [209] X. Fang, J. A. Blazek, J. E. McEwen, and C. M. Hirata, *JCAP* **2017**, 030 (2017), arXiv:1609.05978 [astro-ph.CO].
- [210] M. Kilbinger et al., *Astron. Astrophys.* **497**, 677 (2009), arXiv:0810.5129 [astro-ph].
- [211] L. Casarini, S. A. Bonometto, E. Tesserotto, and P. S. Corasaniti, *JCAP* **1608**, 008 (2016), arXiv:1601.07230 [astro-ph.CO].
- [212] R. Trotta (2017) arXiv:1701.01467 [astro-ph.CO].
- [213] V. Elvira, L. Martino, and C. P. Robert, *International Statistical Review* (2022), 10.1111/insr.12500.
- [214] P. Marshall, N. Rajguru, and A. Slosar, *Phys. Rev.* **D73**, 067302 (2006), arXiv:astro-ph/0412535 [astro-ph].
- [215] H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford, Oxford, England, 1961).
- [216] W. Handley and P. Lemos, *Phys. Rev. D* **100**, 043504 (2019), arXiv:1902.04029 [astro-ph.CO].
- [217] P. Lemos, F. Köhlinger, W. Handley, B. Joachimi, L. White-way, and O. Lahav, *Mon. Not. Roy. Astron. Soc.* **496**, 4647 (2020), arXiv:1910.07820 [astro-ph.CO].
- [218] S. Kullback and R. A. Leibler, *Ann. Math. Statist.* **22**, 79 (1951).
- [219] W. Handley and P. Lemos, *Phys. Rev. D* **100**, 023512 (2019), arXiv:1903.06682 [astro-ph.CO].
- [220] M. Raveri and C. Doux, *Phys. Rev. D* **104**, 043504 (2021), arXiv:2105.03324 [astro-ph.CO].
- [221] M. Raveri and W. Hu, *Phys. Rev. D* **99**, 043506 (2019), arXiv:1806.04649 [astro-ph.CO].
- [222] Y. Park and E. Rozo, *Mon. Not. Roy. Astron. Soc.* **499**, 4638 (2020), arXiv:1907.05798 [astro-ph.CO].
- [223] C. G. Bryden, *J. Inst. Math. Appl.*, 76–90 (1970).
- [224] R. Fletcher, *Comp. J.*, 317–322 (1970).
- [225] D. F. Goldfarb, *Math. Comp.*, 23–26 (1970).
- [226] D. Shanno, *Math. Comp.*, 647–656 (1970).
- [227] H. Akaike, *IEEE T. Automat. Contr.*, 716 (1974).
- [228] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, *Journal of the Royal Statistical Society Series B* **64**, 583 (2002).
- [229] T. Tröster et al. (KiDS), *Astron. Astrophys.* **649**, A88 (2021), arXiv:2010.16416 [astro-ph.CO].
- [230] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., Chapman & Hall/CRC Texts in Statistical Science (Chapman & Hall/CRC, Philadelphia, PA, 2013).

#### Appendix A: HALOFIT validation for $w_0 - w_a$

We model the matter power spectrum by using CAMB [198, 199] to compute the linear matter power spectrum and the HALOFIT semi-analytic fitting formula from Ref. [81] (with the prescription from Ref. [82] for massive neutrinos) to compute non-linear corrections. The HALOFIT fitting formula depends on the linear matter power

spectrum and a subset of cosmology parameters, and was developed based on fits to  $w$ CDM  $N$ -body simulations. While those simulations include cosmological models with  $w \neq 1$ , they do not include cases where the dark energy equation of state varies with time.

To validate the use of HALOFIT to model the non-linear matter power spectrum in our  $w_0 - w_a$  model, we note that Casarini et al. [211] provides a recipe for computing the non-linear power spectrum  $P(k, z)$  for  $w_0 - w_a$  model given modeling ingredients for  $w$ CDM. That scheme works by identifying an effective  $w$ CDM model for each redshift  $z$  for a given  $w_0 - w_a$  cosmology, chosen so  $w_{\text{eff}}(z)$  matches the distance from redshift  $z$  to that of last scattering. Using  $N$ -body simulations, Casarini et al. [211] shows that this mapping can be used to accurately compute nonlinear matter power spectra in  $w_0 - w_a$  cosmologies. While employing the Casarini mapping is not practical for our full analysis because it is too computationally intensive, we can use it for validation. To do this, we use the Casarini prescription to compute the non-linear matter power spectrum for a grid of cosmologies in spanning our two-dimensional  $w_0 - w_a$  prior range. We then compare predictions for the  $3 \times 2$ pt data using our fiducial pipeline to those using the Casarini nonlinear power spectra. For each  $w_0 - w_a$  gridpoint, we evaluate  $\Delta\chi^2$  between our fiducial and the Casarini model predictions and show the results in Fig. 17. Differences between these calculations have  $\Delta\chi^2 < 0.24$  for all allowed values of  $w_0$  and  $w_a$ , with the largest differences occurring in the high- $w_0$ , low- $w_a$  part of parameter space. For all regions where  $w_a > -2.5$ , this difference is  $\Delta\chi^2 < 0.15$ . Based on these results, we conclude that our fiducial model using HALOFIT is accurate enough to perform our  $w_0 - w_a$  analysis with our fiducial scale cuts.

We posit that a significant driver of the Casarini  $w_{\text{eff}}(z)$  mapping’s success in modeling non-linear power for  $w_0 - w_a$  comes from the fact that it correctly accounts for the impact of dynamic dark energy on the linear growth factor. Thus, the fact that we are correctly computing the linear matter power spectrum for the  $w_0 - w_a$  cosmology allows us to reach this accuracy even if we are not explicitly accounting for the  $w_a$  parameter in HALOFIT.

#### Appendix B: Validation of Weyl potential pipeline

As an additional validation of the modeling pipeline used to constrain the modified gravity  $\Sigma_0 - \mu_0$  model, we compare parameter estimates in  $\Lambda$ CDM for the baseline COSMOSIS pipeline used in DES Y3  $3 \times 2$ pt analyses to the pipeline modified to use the Weyl potential to model lensing-related quantities (see Sec. III E for more details). Figure 18 shows  $\Lambda$ CDM constraints obtained by analyzing  $3 \times 2$ pt data using the same linear scale cuts applied for the  $\Sigma_0 - \mu_0$  analyses. That figure shows the results for the baseline pipeline in black and the Weyl pipeline in blue. Differences are negligible for the estimated cosmological parameters  $\Omega_m$  and  $S_8$ .

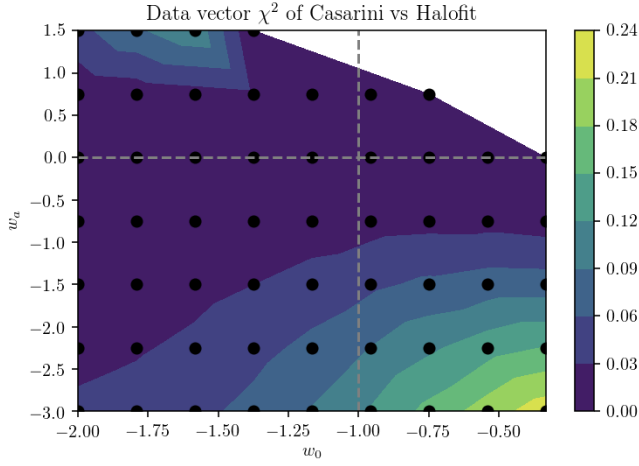


FIG. 17. Validation of HALOFIT for nonlinear modeling in  $w_0 - w_a$ , showing the  $\Delta\chi^2$  difference between the  $3 \times 2$ pt data vector computed with the fiducial pipeline versus one where the Casarini [211] method has been used to compute the nonlinear matter power spectrum. The colored heatmap shows the interpolation between points sampled in a grid (black circles) in the  $w_0 - w_a$  plane. The white region in the upper right reflects the excluded region where  $w_0 + w_a > 0$ .  $\Delta\chi^2 \ll 1$  across the full space, indicating that the modeling of HALOFIT is sufficient.

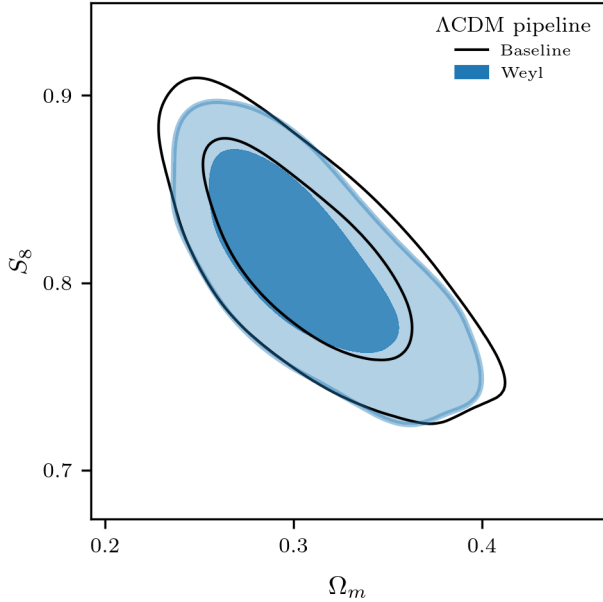


FIG. 18. Constraints on  $\Omega_m$  and  $S_8$  in  $\Lambda$ CDM, showing 68 and 95% confidence limits from the baseline DES Y3 COSMOSIS pipeline in black and the modified version using the Weyl potential for the lensing predictions in blue. These results are obtained using  $3 \times 2$ pt measurements for the same linear scale cuts used to obtain  $\Sigma_0 - \mu_0$  constraints.

## Appendix C: The FASTISMORE framework for robustness tests

The FASTISMORE framework used for the model robustness tests described in Sec. VB is at its core an application of importance sampling for performing validation tests of the robustness against systematics. Importance sampling (IS) is a method to quickly estimate a target distribution  $p(\Theta)$  by using samples from a proposal distribution  $q(\Theta)$  reweighted by the ratio  $p(\Theta)/q(\Theta) \propto \mathcal{L}_p/\mathcal{L}_q$  (see e.g. Ref. [212]). The FASTISMORE framework consists of several parts: the infrastructure for performing fast IS posterior estimates, code for computing quality statistics for that estimate as well as for how the IS procedure contributes to sampling variance, and guidelines for using those metrics based on a number of simulated analyses. These tools will be made publicly available and documented in more detail in an upcoming publication, Ref. [154].

The accuracy of an IS estimate depends on whether the samples drawn from the proposal distribution cover the relevant parameter space of the target distribution with high enough density. In our application for robustness tests, the proposal distribution is the posterior of our baseline simulated analysis and the target is the posterior for a synthetic data vector contaminated with a systematic that is not included in the analysis model, as described in Sec. VB. Given this, chains will only be required if a contamination produces a significant change in the posterior. For cases where IS can be used, this method allows the impact of a systematic to be assessed in seconds, as opposed to the thousands of core-hours required to run a chain.

We quantify the performance of IS posterior estimates using Kish's effective sample size:

$$\text{ESS} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}, \quad (\text{C1})$$

where  $w_i$  is the total weight of each sample. Assuming this approximates the true effective sample size (for a discussion on the validity of this approximation, see [213]), one can estimate the standard error on the mean  $\bar{\theta}$  of the parameter of interest  $\theta$  by using  $\sigma_{\bar{\theta}} \approx \sigma_{\theta}/\sqrt{\text{ESS}}$ , where  $\sigma_{\theta}$  is the standard deviation of the posterior. The uncertainty on the parameter mean shift is then estimated as:

$$\sigma_{\Delta\bar{\theta}} \lesssim \sigma_{\theta} \sqrt{\frac{1}{\text{ESS}_{\text{base}}} + \frac{1}{\text{ESS}_{\text{cont}}}}, \quad (\text{C2})$$

where  $\text{ESS}_{\text{base}}$  is the effective sample size of the baseline chain and  $\text{ESS}_{\text{cont}}$  is the effective sample size of the IS-estimated contaminated samples. For the analysis of contaminated synthetic data described in VB of this paper, we found  $\text{ESS}_{\text{cont}} > 150$  for all validation tests such that the uncertainty on the parameter mean shifts are at most  $0.08\sigma$ . Thus all of our systematics tests pass the IS quality requirements.

We also employ the FASTISMORE framework to assess the impact of systematics on model comparison metrics between  $\Lambda$ CDM and a beyond- $\Lambda$ CDM models. To assess how Suspiciousness (see Appendix F)  $\ln S$  is affected by contamination from systematics, we must compare quantities derived from

four chains: a baseline ( $B$ ) and alternative (contaminated) synthetic data ( $A$ ) chain for both the extended model ( $X$ ) and  $\Lambda$ CDM ( $0$ ). As will be shown in Ref. [154], the change in  $\ln S$  due to systematic contamination can be written in terms of the within-model differences between the baseline ( $B$ ) and alternative ( $A$ ) synthetic data,

$$\Delta \ln S_{X0} = \ln S_{X0}^A - \ln S_{X0}^B \quad (C3)$$

$$= \ln S_{AB}^X - \ln S_{AB}^0. \quad (C4)$$

Here, we use  $S_{ab}^c$  to denote Suspiciousness between  $a$  and  $b$ , keeping  $c$  fixed, defined so  $\ln S < 0$  indicates a preference for  $a$ . This rearrangement makes the calculation more tractable, as the quantities  $S_{AB}^X$  and  $S_{AB}^0$  are easily computed using importance sampling.

#### Appendix D: Additional validation plots and discussion

Here we provide additional information to supplement the validation test results presented in Sec. V.

##### 1. All-data robustness plots

We begin by showing plots complementing Fig. 1 of Sec. V B and Fig. 3 of Sec. V C, which show the impact of data vector contamination and model variations on the combined  $3 \times 2$ pt+BAO+RSD+SN+*Planck* constraints. Since those all-data constraints are much more constraining than the DES  $3 \times 2$ pt alone for some models, the effects of data vector contamination and model variations on the combined constraints are not clearly visible in plots in the main body of the text, whose ranges are set to show the  $3 \times 2$ pt-only constraints. To facilitate closer examination, here we include versions of those plots showing only the all-data constraints. Fig. 19 shows an All-data only version simulated analysis plot, while Fig. 20 shows the real data response to model variations.

In both plots, we additionally show the results of the robustness tests for  $N_{\text{eff}} - m_{\text{eff}}$  when the priors requiring  $\Delta N_{\text{eff}} > 0.047$  or  $m_{\text{th}} < 10$  eV are applied. We set the  $m_{\text{eff}}$  axis range to show those results clearly, cutting off the fiducial prior points which were shown in Sec. V C to have non-negligible shifts. For these alternative  $N_{\text{eff}} - m_{\text{eff}}$  priors, the nonlinear bias and baryon contamination produces a shift  $\Delta m_{\text{eff}} = -0.33$ , which is only slightly above and within sampling variance of the 0.3 threshold.

##### 2. Robustness investigation: $N_{\text{eff}} - m_{\text{eff}}$

In the robustness tests of Sec. V C, we noted that  $N_{\text{eff}} - m_{\text{eff}}$  parameter constraints shifted significantly in response to several model variations. Here we provide further description of these non-negligible parameter shifts and argue that they are likely the result of prior volume effects associated with an unconstrained region of parameter space at small  $\Delta N_{\text{eff}}$ . For  $N_{\text{eff}} - m_{\text{eff}}$ , recall that we will be primarily focused on

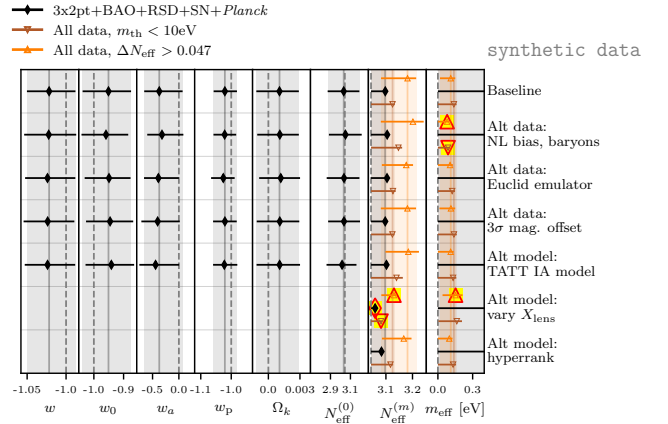


FIG. 19. Constraints on beyond- $\Lambda$ CDM model parameters for the same simulated analyses studied in Fig. 1 of Sec. V B, but with narrower axis ranges for models where All-data is much more constraining than DES  $3 \times 2$ pt. Points and error bars show the mean and 68% confidence interval for marginalized constraints on extended model parameters. Yellow and red highlights indicate shifts larger than  $0.3\sigma$  according to Eq. (31). Vertical dashed lines show the input,  $\Lambda$ CDM parameter values, while solid vertical lines and shaded regions show the location of the baseline results. For the effective number of radiative degrees of freedom,  $N_{\text{eff}}^{(0)}$  is the constraint for the model with no sterile neutrino mass, and  $N_{\text{eff}}^{(m)}$  shows the constraints for the  $N_{\text{eff}} - m_{\text{eff}}$  model. For  $N_{\text{eff}} - m_{\text{eff}}$  we additionally show results using the alternative prior  $\Delta N_{\text{eff}} > 0.047$ , which are less subject to projection effects and thus more robust than posteriors produced using the fiducial  $N_{\text{eff}} - m_{\text{eff}}$  prior.

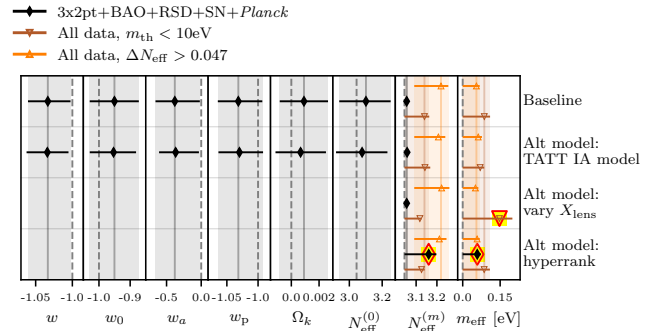


FIG. 20. Real All-data constraints for the same analyses presented in Fig. 3 of Sec. V C, but with narrower axis ranges for parameters where All-data is much more constraining than DES  $3 \times 2$ pt. Points and error bars show the mean and 68% confidence interval for marginalized constraints on extended model parameters, yellow and red highlights indicate shifts larger than  $0.3\sigma$  and shaded regions indicate the location of the baseline results.

results from all data (DES  $3 \times 2$ pt+BAO+RSD+SN+*Planck*) rather than DES  $3 \times 2$ pt only results.

Of the parameter shifts that are above our desired threshold for  $N_{\text{eff}} - m_{\text{eff}}$ , those produced by the TATT model variation are the least significant, and are not very concerning. For the simulated analysis, switching to TATT causes  $\Delta m_{\text{eff}} = 0.37$ . This change in  $m_{\text{eff}}$  is not concerning, both because it is only

marginally over our threshold, and because our simulated constraints on  $m_{\text{eff}}$  are one-sided upper bounds so that this shift can be interpreted as being simply due to weakened constraining power in a more complicated IA model. When these tests were repeated for real data, shifts in the constraints on both  $N_{\text{eff}}$  and  $m_{\text{eff}}$  were negligible.

The  $X_{\text{Lens}}$  and hyperrank tests have a more dramatic impact on the  $N_{\text{eff}} - m_{\text{eff}}$  constraints from all data. For the analysis of real data, varying  $X_{\text{Lens}}$  causes  $\Delta_{m_{\text{eff}}} = -0.41$ , while using hyperrank causes  $\Delta_{N_{\text{eff}}} = 0.91$  and, strikingly,  $\Delta_{m_{\text{eff}}} = -18$ . To understand this behavior, we note that the upper bound on  $m_{\text{eff}}$  for all data is largely determined by how the posterior is shaped in the low- $\Delta_{N_{\text{eff}}}$  region of parameter space where  $m_{\text{eff}}$  has no impact on observables because the sterile neutrinos are indistinguishable from cold dark matter. Because CMB measurements provide tight upper bounds on  $\Delta_{N_{\text{eff}}}$ , a significant fraction of the all data posterior occupies this region of parameter space. The flatness of the likelihood in the  $m_{\text{eff}}$  direction when  $\Delta_{N_{\text{eff}}}$  is low (independent of what data are considered) implies that the marginalized constraints on both  $N_{\text{eff}}$  and  $m_{\text{eff}}$  are highly susceptible to the details of how the higher-dimensional posterior projects those parameter directions. Given this, our all-data  $N_{\text{eff}} - m_{\text{eff}}$  constraints may be highly sensitive to the choice of nuisance parameters, and the data's noise realization may produce large shifts in  $m_{\text{eff}}$  that do not necessarily carry physical information. As we describe in Appendix D 4, hyperrank causes hard-to-characterize changes to the posterior shape even in  $\Lambda\text{CDM}$ , so it is plausible that the dramatic  $\Delta_{m_{\text{eff}}}$  produced by hyperrank is related to this kind of parameter-space projection.

This prior-volume-effect hypothesis is supported by the fact that our results become more robust when we apply priors to remove the unconstrained low- $\Delta_{N_{\text{eff}}}$  region. We consider two such priors. We run additional (real data) chains raising the lower bound on  $N_{\text{eff}}$  to require  $\Delta_{N_{\text{eff}}} > 0.047$ , and additionally we consider a cut on the physical sterile neutrino mass, assuming a thermal relic model,  $m_{\text{th}} < 10\text{ eV}$ . This  $m_{\text{th}}$  cut matches the prior used for  $N_{\text{eff}} - m_{\text{eff}}$  in the *Planck* 2018 analysis [11]. Plots showing these shifts for all-data constraints with the alternative  $\Delta_{N_{\text{eff}}}$  priors can be found above in Figs. 19 and 20.

Requiring  $\Delta_{N_{\text{eff}}} > 0.047$  leads to negligible shifts between the all-data constraints from the hyperrank and the baseline models, as well as between the baseline and varying  $X_{\text{Lens}}$ . Requiring  $m_{\text{th}} < 10\text{ eV}$ , which is the prior matching the *Planck* 2018 analysis, results in negligible shifts due to  $X_{\text{Lens}}$ , and hyperrank-vs-baseline shifts of  $\Delta_{(N_{\text{eff}}, m_{\text{eff}})} = (-0.26, -0.39)$ , only slightly above our  $0.3\sigma$  threshold. For comparison, we also consider simulated  $N_{\text{eff}} - m_{\text{eff}}$  analyses of all data combined, post-processing the chains to enforce the alternative priors rather than running new chains. For simulated all-data analyses with fiducial priors,  $X_{\text{Lens}}$  produces  $\Delta_{(N_{\text{eff}}, m_{\text{eff}})} = (-0.71, +0.57)$  and hyperrank produces  $(-0.27, +0.39)$ , while with the alternate priors hyperrank causes negligible shifts and  $X_{\text{Lens}}$  produces a shift of  $\Delta_{N_{\text{eff}}} \sim 0.5$ . While this latter shift is non-negligible, it is not concerning given the robustness of the real data results, especially noting that the post-processing used for these synthetic

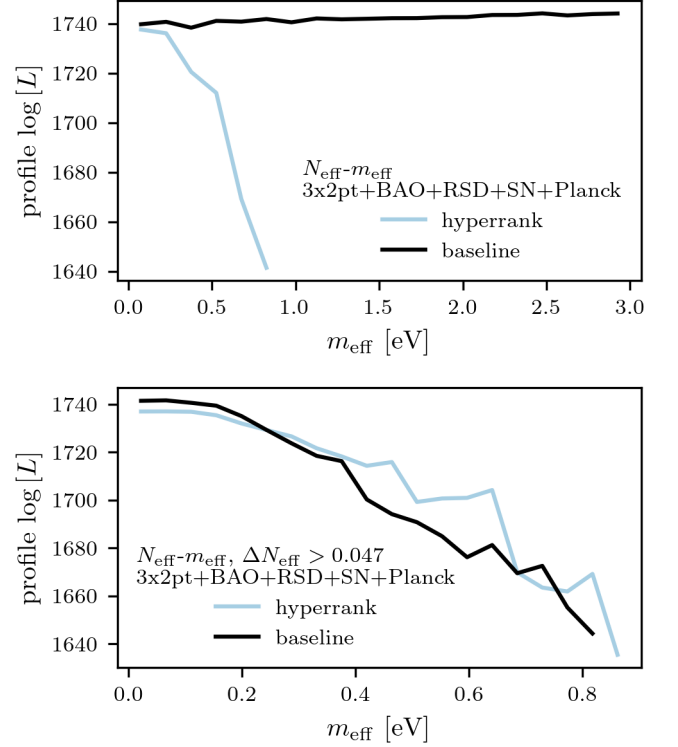


FIG. 21. Profile likelihood for All-data  $N_{\text{eff}} - m_{\text{eff}}$  chains, comparing of baseline and hyperrank. Lines show the maximum likelihood found for a chain sample in each of 20 bins of  $m_{\text{eff}}$  values. The upper panel shows results for the fiducial  $N_{\text{eff}} - m_{\text{eff}}$  priors, and the lower panel shows results for chains run with the alternative  $\Delta_{N_{\text{eff}}} > 0.047$  prior. The fact that hyperrank does not produce a better fit to the data than the baseline at small  $m_{\text{eff}}$  supports the idea that shift in  $m_{\text{eff}}$  constraints for hyperrank is due to parameter space projection effects rather than an actual strong preference for small  $m_{\text{eff}}$  values.

data tests results in significant sampling uncertainty.

To lend further support to the idea that the dramatic shift in the all-data  $N_{\text{eff}} - m_{\text{eff}}$  constraints (particularly  $m_{\text{eff}}$ ) when using hyperrank is a prior volume effect, we study the profile likelihood for  $m_{\text{eff}}$  in Fig. 21. The profile likelihood shows the maximum likelihood in 20 bins of sampled  $m_{\text{eff}}$  values. For the fiducial  $N_{\text{eff}} - m_{\text{eff}}$  priors, we see in the top panel that while the hyperrank maximum sampled likelihood drops at large  $m_{\text{eff}}$ , the hyperrank model does not produce a better fit to the data than the baseline chain in the small  $m_{\text{eff}}$  regime. The lower panel shows the same profile for chains run with a prior requiring  $\Delta_{N_{\text{eff}}} > 0.047$ . In that panel, the hyperrank and baseline profile likelihoods remain fairly similar as  $m_{\text{eff}}$  increases. This suggests that that drop seen for the hyperrank chain in the top panel is due to a lack of chain samples exploring the high  $m_{\text{eff}}$ , rather than a dramatically worse fit. This supports the idea that the differences between the marginalized hyperrank and baseline posteriors are driven by parameter space projection effects.

Overall these studies support the conclusion that our  $N_{\text{eff}} - m_{\text{eff}}$  constraints are significantly more stable against model

changes when a prior is applied to remove the unconstrained small- $\Delta N_{\text{eff}}$  region. This motivates our choice to focus on results reported for  $N_{\text{eff}} - m_{\text{eff}}$  constraints obtained with the  $\Delta N_{\text{eff}} > 0.047$  and  $m_{\text{th}} < 10\text{eV}$  priors rather than those obtained with the fiducial prior.

### 3. Robustness investigation: binned $\sigma_8(z)$

Here we describe and investigate non-negligible shifts in the binned  $\sigma_8(z)$  parameters due to the hyperrank and  $X_{\text{Lens}}$  model variations studied in Sec. V C. The only  $> 0.3\sigma$  shift produced by varying  $X_{\text{Lens}}$ ,  $\Delta_{A_3^{P_{\text{lin}}}} = 0.36$  for DES  $3\times 2\text{pt}$ -alone, is not far above our threshold so is not concerning. We thus focus primarily on the impact of the hyperrank approach to parameterizing source galaxy redshift uncertainties. For the real-data  $3\times 2\text{pt}$ -only analysis, using hyperrank leads to parameter shifts of  $\Delta_{(A_2^{P_{\text{lin}}}, A_3^{P_{\text{lin}}})} = (+0.72, +0.84)$ , while analyzing all data produces even more significant parameter shifts:  $\Delta_{(A_2^{P_{\text{lin}}}, A_3^{P_{\text{lin}}}, A_4^{P_{\text{lin}}})} = (+1.73, +1.36, +1.42)$ .

A detailed characterization of what is driving the low-redshift amplitude shifts is beyond the scope of this paper, though we present some exploratory investigation in Appendix D 4. What these findings clearly highlight is that source photo- $z$  uncertainties and our method of accounting for them can have a significant impact on inferences about the growth of LSS over time, especially when constraining models that, like our binned  $\sigma_8(z)$  parameterization, add degrees of freedom beyond what is expected for  $\Lambda\text{CDM}$ . While in principle the hyperrank method should provide a more complete description of photo- $z$  uncertainties than the baseline  $\Delta z_s$  nuisance parameters, without additional validation we are not confident in switching our main analysis to use it for binned  $\sigma_8(z)$ . Given this ambiguity, we choose to report for binned  $\sigma_8(z)$  constraints and model comparisons for both our baseline model and for hyperrank. Showing results from both analyses will roughly quantify the size of the photo- $z$ -related systematic uncertainties. Note that this means that if we find tension with  $\Lambda\text{CDM}$  in one but not both of these binned  $\sigma_8(z)$  analyses, we will not be able to definitively claim a discovery of non-standard large scale structure growth.

To further characterize what we can or cannot say robustly about our binned  $\sigma_8(z)$  inferences, we repeat these validation tests<sup>22</sup> for the derived parameters  $\sigma_8^{[\text{bin } i]}$  defined in Eq. (29). Recall that  $\sigma_8^{[\text{bin } i]}$  are more closely related to the amplitude of large scale structured observed separately in each redshift bin than the sampled  $A_i^{P_{\text{lin}}}$  amplitudes, which are defined relative to the amplitude in the lowest redshift bin. When real data are analyzed, model variations still produce non-negligible  $\sigma_8^{[\text{bin } i]}$  shifts, as can be seen in Fig. 22, but these are smaller than those found for the sampled  $A_i^{P_{\text{lin}}}$  parameters. The shift for the lowest redshift bin shifts in response to

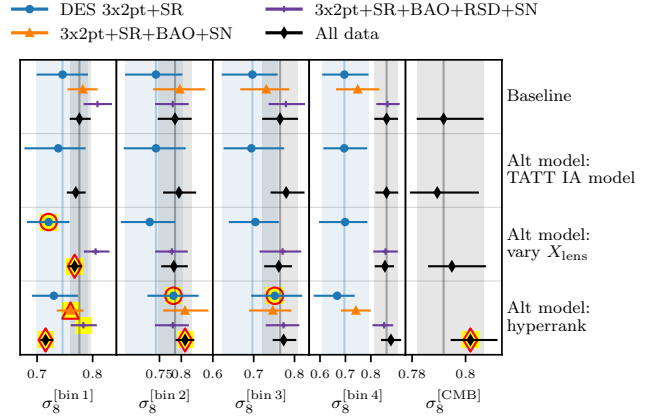


FIG. 22. Impact of model variations on the derived parameters  $\sigma_8^{[\text{bin } i]} \equiv \sigma_8[A_i^{P_{\text{lin}}}]^{1/2}$ . Plot formatting matches that of Fig. 3 and All data (in black) refers to DES  $3\times 2\text{pt}+\text{SR}+\text{BAO}+\text{RSD}+\text{SN}+\text{Planck}$ .

varying  $X_{\text{Lens}}$  are only slightly above the  $0.3\sigma$  threshold, with  $\Delta_{\sigma_8^{[\text{bin } 1]}} = -0.42$  and  $-0.39$  for DES  $3\times 2\text{pt}$  and all data respectively. For hyperrank, the non-negligible shifts for  $3\times 2\text{pt}$ -only are  $\Delta_{(\sigma_8^{[\text{bin } 2]}, \sigma_8^{[\text{bin } 3]})} = (+0.47, +0.64)$ , while for all data they are  $\Delta_{(\sigma_8^{[\text{bin } 1]}, \sigma_8^{[\text{bin } 2]}, \sigma_8^{[\text{CMB}]})} = (-2.72, +0.52, +0.60)$ . For  $3\times 2\text{pt}+\text{BAO}+\text{RSD}+\text{SN}$  (leaving out *Planck*), the only non-negligible shift is  $\Delta_{\sigma_8^{[\text{bin } 1]}} = -0.73$ , and similarly  $3\times 2\text{pt}+\text{BAO}+\text{SN}$  (DES combined with only geometric external data) it is  $\Delta_{\sigma_8^{[\text{bin } 1]}} = -0.62$ .

Thus, we see that when we combine  $3\times 2\text{pt}$  with external constraints, the source photo- $z$  marginalization scheme primarily contributes systematic uncertainty to the LSS amplitude measured for the lowest redshift bin. The fact that the sampled  $A_i^{P_{\text{lin}}}$  parameters are defined relative to bin 1 is thus why those amplitudes are strongly affected by the change to hyperrank. Our study of the  $\sigma_8^{[\text{bin } i]}$  derived parameters show that when we combine DES  $3\times 2\text{pt}$  measurements with other low redshift geometric probes, we are in fact able to make fairly robust inferences of binned  $\sigma_8(z)$  for the redshift ranges corresponding to bins 2-4.

### 4. Hyperrank discussion

In our model robustness validation tests, for both  $N_{\text{eff}} - m_{\text{eff}}$  and for binned  $\sigma_8(z)$  we found that using the hyperrank [155] method to marginalize over source photo- $z$  uncertainties, as opposed to the fiducial  $\Delta z_s$  mean-shift nuisance parameters, caused non-negligible shifts in the beyond- $\Lambda\text{CDM}$  parameter posteriors. Here we present some additional investigation into that behavior. To place these studies in context, we illustrate that switching to hyperrank can have non-negligible impacts on parameter estimation even in  $\Lambda\text{CDM}$ . The method's impact on shear-only analysis has been thoroughly studied and the DES Y3 cosmic shear results were

<sup>22</sup> Note that these redefined amplitude parameters, and thus these additional robustness tests, were explored after unblinding.

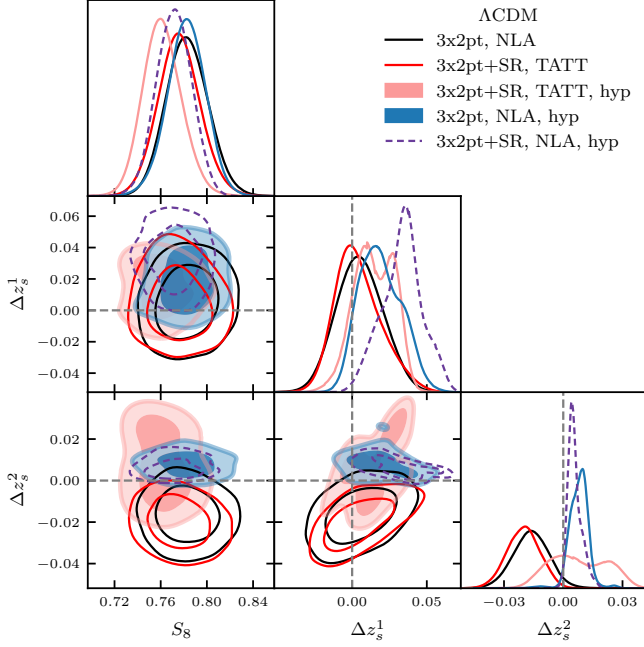


FIG. 23. Comparison of baseline and hyperrank chain posteriors for  $S_8$  and shifts in mean source bin redshift around the fiducial  $n_s(z)$  means for bins 1 and 2. This is a  $3 \times 2$ pt version of the plot shown for shear-only in Fig. 17 of Ref. [67].

found to be robust to this model variation [67]. However, in DES-Y3KP it was found that switching to hyperrank produces a  $0.53\sigma$  shift in  $S_8$ .

Fig. 23 further illustrates this behavior by showing the  $\Lambda$ CDM constraints on  $S_8$  and the mean redshift of a subset of source bins for various iterations of our analysis choices. In that figure, solid red and black lines show  $\Lambda$ CDM posteriors on the photo- $z$  bias nuisance parameters for the baseline settings in this work (black, using NLA as the IA model and not including shear ratio) and in the  $\Lambda$ CDM analysis of DES-Y3KP (red, TATT IA model, including shear ratio). The shaded pink contours show the results for the hyperrank chain that was run as part of robustness tests in DES-Y3KP (with TATT and shear ratio), while dashed purple contours and shaded blue contours show hyperrank chains run as part of this paper’s beyond- $\Lambda$ CDM studies (with NLA), with and without including the shear ratio likelihood, respectively. We see that all of the hyperrank chains have multi-modal posteriors, and that there are significant qualitative difference between the various hyperrank chains compared. The choice of IA model has the largest impact on the shape of these posteriors, but even comparing the two NLA chains (blue-shaded and purple-dashed), including or not including shear ratio can also cause non-negligible posterior shifts. Notably, we find that these model variations have a more significant impact on  $S_8$  estimates when hyperrank is used than in the baseline  $\Delta z$  approach to marginalizing photo- $z$  uncertainties.

These effects can be understood in terms of interactions between intrinsic alignment (IA) parameters and details of the shape of the source redshift distribution. The contribution of

IA to measured cosmic shear become fractionally more important at low source galaxy redshifts where lensing contributions are smaller. Additionally, the IA signal depends on the projection of  $n_s^2(z)$  onto the sky, while lensing depends on the square of the projected  $n_s(z)$ . Together, these two effects mean that IA calculations are more sensitive to the detailed shape of the source galaxy redshift distribution, and thus that different hyperrank  $n_s(z)$  realizations can have very different IA kernels, especially depending on their low- $z$  features. While including or not including the shear ratio likelihood doesn’t significantly impact the cosmology constraints in the baseline  $3 \times 2$ pt analysis, this choice does have an impact on the IA parameter constraints, and thus affects the posterior when hyperrank is used.

If this behavior is occurring in  $\Lambda$ CDM, it perhaps not surprising that hyperrank causes significant shifts in the  $A^{P_{\text{lin}}}$  amplitudes when we allow  $\sigma_8$  to vary independently in different redshift bins. We explored whether specific features in the  $n(z)$  distributions sampled by hyperrank correlate with the binned  $\sigma_8(z)$  amplitude parameters, and thus might be driving the shifts we see. Unfortunately, this investigation did not yield any further insights.

#### Appendix E: Metrics for assessing tension between datasets

In Sec. VII, we employ three different tension metrics to assess agreement between datasets: the Bayes ratio  $R$ , Suspiciousness  $S$ , and a  $p$ -value computed from  $S$  and the Bayesian model dimensionality  $d_{BMD}$ . The Bayes ratio  $R$  is defined for independent datasets  $A$  and  $B$  and for their combination  $AB$  as [214]:

$$R \equiv \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A \mathcal{Z}_B}, \quad (\text{E1})$$

where

$$\mathcal{Z}_D \equiv P(\mathbf{D}|\mathcal{M}) = \int d\boldsymbol{\Theta} \mathcal{L}(\mathbf{D}|\boldsymbol{\Theta}, \mathcal{M}) \pi(\boldsymbol{\Theta}|\mathcal{M}). \quad (\text{E2})$$

is the Bayesian Evidence. In that expression,  $\mathcal{L}$  is the likelihood of observing the data given model  $\mathcal{M}$  and parameter values  $\boldsymbol{\Theta}$ , and  $\pi$  is the prior probability of those parameters given the model. In effect  $R$  can be viewed as a hypothesis test assessing the odds of both datasets being described with a single set of parameters ( $\mathcal{Z}_{AB}$ ) as opposed to two independent sets of parameters ( $\mathcal{Z}_A \mathcal{Z}_B$ ). Smaller values of  $R$  indicate stronger evidence of tension between measurements from datasets  $A$  and  $B$ . When using  $\ln R$  the strength of the tension is usually interpreted using the Jeffreys’ scale [215], where  $\ln R < -2.3$  is considered ‘strong’ tension with approximately 10 : 1 odds,  $-2.3 < \ln R < -1.2$  is considered ‘substantial’ tension with approximately 3 : 1 odds, and  $\ln R > -1.2$  indicates the datasets are in agreement. This interpretation of odds is only correct in the context where one of the models being considered is correct and where the priors accurately characterize prior beliefs on the parameters. As is discussed in e.g. Refs. [158, 216, 217], the value of  $\ln R$  depends strongly on the choice of parameter prior ranges, so

when interpreting tension assessed with the Bayes ratio, one should check the robustness of conclusions under reasonable changes to those priors. The fact that we use wide, uninformative priors therefore makes  $\ln R$  somewhat ambiguous to interpret as a tension metric for this work.

Given this, we additionally report tension using the Bayesian Suspiciousness  $S$  [216]. Like the Bayes ratio, Suspiciousness measures tension between posteriors in our full sampled parameter space, but removes the prior dependence by dividing  $R$  by the information ratio  $I$ ,

$$\ln S = \ln R - \ln I. \quad (\text{E3})$$

The information ratio quantifies the probability of data  $A$  and  $B$  given the prior width, and is defined

$$\ln I \equiv \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB}, \quad (\text{E4})$$

where  $\mathcal{D} \equiv \int \mathcal{P} \ln(\mathcal{P}/\pi) d\theta$  is the Kullback–Leibler Divergence [218] from the prior  $\pi$  to the posterior  $\mathcal{P}$  for a given dataset. One can interpret the Suspiciousness as a posterior-averaged goodness-of-fit statistic between the combined vs. independent datasets  $A$  and  $B$  [19].

As with the Bayes ratio, more negative values of  $\ln S$  indicate stronger evidence of tension, while positive values indicate agreement between datasets. To further quantify the strength of tension or agreement, we can use the fact that if both datasets come from the same set of parameters and there is some choice of parameters in which the posterior is roughly Gaussian, the quantity  $d_{\text{BMD}} - 2 \ln S$  follows approximately a  $\chi^2_{d_{\text{BMD}}}$  probability distribution, where  $d_{\text{BMD}}$  counts the number of parameter dimensions constrained by both posteriors  $A$  and  $B$  [216]. We use this information to compute a  $p$ -value estimating the probability of finding a value of  $\ln S$  as small or smaller than the measured value if the two datasets are actually in agreement.

For each chain we assess  $d_{\text{BMD}}$  using the Bayesian Model Dimensionality [219], which estimates the number of parameters constrained by a given posterior. It is equal to

$$d = 2 \left( \langle (\ln \mathcal{L})^2 \rangle - \langle \ln \mathcal{L} \rangle^2 \right), \quad (\text{E5})$$

where again angled brackets refer to the posterior-weighted average. We compute the number of parameters that are independently constrained by both datasets  $A$  and  $B$  as

$$d_{\text{BMD}} = d_A + d_B - d_{AB}, \quad (\text{E6})$$

where the  $d_{AB}$  term avoids double-counting parameters. We evaluate the survival function,

$$p(S, d_{\text{BMD}}) = \int_{d_{\text{BMD}} - 2 \ln S}^{\infty} \chi^2_{d_{\text{BMD}}}(x) dx \quad (\text{E7})$$

There are several other metrics one could use to assess tensions between datasets, including but not limited to parameter difference distributions [220, 221] and eigentensions [222]. It was shown in Ref. [158] that all of these metrics give results in agreement with  $\ln S$  when quantifying tensions. We therefore focus on  $S$ , which has the advantage of requiring no additional computation after running the chains.

## Appendix F: Model comparison metrics

In Sec. VIJ we perform a number of model comparison tests to assess whether data favor any extensions to  $\Lambda$ CDM over that cosmological standard model. Here we present definitions of the metrics used to assess those comparisons, divided into two categories: Bayesian quantities depending on the posterior distribution in the full parameter space, and those depending on maximum *a posteriori* probability (MAP) estimates. To define these metrics, let us consider two models: a baseline model  $\mathcal{M}_0$  and a model which extends it,  $\mathcal{M}_X$ , such that the parameter space of  $\mathcal{M}_0$  is a subspace of  $\mathcal{M}_X$ 's parameter space.

### 1. Bayesian quantities

The Bayesian quantities are analogous to those used to evaluate tensions between datasets. We define the Bayes ratio by comparing the Bayesian evidence evaluated in these different models but with matching data via,

$$R = \frac{\mathcal{Z}_0}{\mathcal{Z}_X}. \quad (\text{F1})$$

Suspiciousness  $S$  and an associated  $p(S, d_{\text{BMD}})$  are defined in the similar way to Eqs. (E3)–E7, but with  $\mathcal{M}_0$  replacing all quantities associated with the  $AB$  joint constraints, and  $\mathcal{M}_X$  replacing the sums where contributions from datasets  $A$  and  $B$  are included separately. That is to say, the definition of  $S$  in Eq. (E3) still holds, but with a single dataset used in all parts of the evaluation and

$$\ln I \equiv \mathcal{D}_X - \mathcal{D}_0, \quad (\text{F2})$$

and similarly,

$$d_{\text{BMD}} \equiv d_X - d_0. \quad (\text{F3})$$

One challenge of using  $\ln S$  for model comparison is that number of additional constrained parameters  $d_{\text{BMD}}$  is typically small, such that noise in the estimate of  $d_{\text{BMD}}$  can have a large impact on  $p(S, d_{\text{BMD}})$ . Specifically, if we let  $\Delta k$  be the number of parameters that are in  $\mathcal{M}_X$  but not in  $\mathcal{M}_0$ , then  $d_{\text{BMD}} \leq \Delta k$ . Most of the models we consider have  $\Delta k = 1$  or  $2$ , meaning that if the additional extended model  $X$  parameters are not well constrained,  $d_{\text{BMD}}$  can be very small. If we could make a noiseless estimate of  $d_{\text{BMD}}$ , this quantity would approach zero in the limit of completely unconstrained beyond- $\Lambda$ CDM parameters. In practice, sampling variance can cause the estimates made in Eqs. (E5) and E6 to return values of  $d_{\text{BMD}} < 0$ . This complicates interpretation of  $p(S, d_{\text{BMD}})$  (Eq. (E7)), because the  $\chi^2$  probability distribution is undefined for negative numbers of degrees of freedom. For estimates where this happens we report  $p(S, d_{\text{BMD}}) = 1$ , reasoning that if the added parameters of  $\mathcal{M}_X$  are unconstrained, no tension with  $\mathcal{M}_0$  can be found.

Given the ambiguity of this determination, we additionally report a tension probability using the number of added parameters  $\Delta k$  instead of  $d_{\text{BMD}}$ :  $p(S, \Delta k)$ . This ensures we

avoid scenarios where the probability distribution for  $S$  is undefined. In addition to being less subject to sampling error,  $p(S, \Delta k)$  has a benefit for interpretation, since it can be viewed as a Bayesian likelihood ratio test. As will be shown in Ref. [154], we can interpret the model-comparison formulation of Suspiciousness as the change in *posterior-averaged* goodness-of-fit,  $2 \ln S = \langle \chi_X^2 \rangle - \langle \chi_0^2 \rangle$ . This means  $p(S, \Delta k)$ , which evaluates this change relative to the expected improvement from additional model freedom, serves as a Bayesian analog of the more traditional  $\chi^2$  test statistic that compares the goodness-of-fit at the two models' maximum likelihood points. Further connections can be made to more traditional information criteria-based model comparison statistics, where for limiting cases the Suspiciousness can be interpreted as analogous to  $\Delta \text{DIC}$ , but with a lesser penalty applied for additional model parameters. This is explored in greater detail in Ref. [154].

## 2. MAP-based statistics

In our large parameter space, MAP estimates from nested sampler chains are subject to significant sampling error and so cannot be accurately determined by simply selecting the sample with the highest posterior from a chain. To estimate the maximum posterior, we therefore perform additional maximization as follows. For each chain, we select the 15 samples with the highest reported posteriors. Starting at each of those 15 points, we run two iterative optimization searches to maximize the posterior using the COSMOSIS `maxlike` sampler, which is an interface to the `scipy.optimize` function [192], using the BFGS [223–226] optimization algorithm. Of the resulting 15 MAP estimates, we select the one with the highest posterior. Based on limited studies for simulated analyses we find this produces reasonably accurate estimates of the maximum posterior probability, (the error on  $\chi^2$  is probably less than about 0.5, though we have not quantified this rigorously), but still very noisy estimates of the associated parameter values. Given this, we use the maximum posterior estimates for model comparison statistics, but we do not report MAP parameter values.

We use the MAP posterior estimates to compute the model comparison statistics  $\Delta \chi^2$ ,  $\Delta \text{AIC}$ , and  $\Delta \text{DIC}$ . The quantity  $\chi^2$  here measures the goodness of fit at the best-fit point in parameter space,

$$\chi^2 = -2 \ln \mathcal{L}^{\max} \quad (\text{F4})$$

In practice we focus on the quantity  $\Delta \chi^2 / \Delta k$ , where  $\Delta \chi^2 = \chi_X^2 - \chi_0^2$  and  $\Delta k$  is the change in modeling degrees of freedom between models  $\mathcal{M}_X$  and  $\mathcal{M}_0$ . The Akaike Information Criterion (AIC) is defined as [227]

$$\text{AIC} = -2 \ln \mathcal{L}^{\max} + 2k \quad (\text{F5})$$

where  $k$  is the number of model parameters. The Deviance Information Criterion (DIC) is defined by [228]

$$\text{DIC} = -2 \ln \mathcal{L}^{\max} + 2p_{\text{DIC}}. \quad (\text{F6})$$

$$\text{with } p_{\text{DIC}} = 2 \ln \mathcal{L}^{\max} - 2 \langle \ln \mathcal{L} \rangle \quad (\text{F7})$$

Note that here we follow Ref. [229] in using a MAP-based calculation of the DIC statistic, rather than an alternative definition where  $p_{\text{DIC}}$  is instead equal to the Bayesian model dimensionality defined in Eq. (E5) (see e.g. Eq. (7.10) in Ref. [230]). This is motivated by the fact that the MAP-based calculation is not affected by the instabilities for small  $d_{\text{BMD}}$  described in the text below Eq. (F3).

The numbers reported for MAP-based quantities in Sec. VIJ rely on a slightly different definitions than those given above. This is because we would like to consider agreement with the Gaussian priors on DES nuisance parameters in addition to the  $3 \times 2\text{pt}$  and external likelihoods to assess goodness-of-fit. For example, if model  $\mathcal{M}_X$  could get an excellent fit to the  $3 \times 2\text{pt}$  measurements, but as a consequence required shear calibration parameters  $m_i$  to take extreme values compared to the principled priors made as part of the METACALIBRATION analysis, we would not want to consider that model favored over  $\mathcal{M}_0$ . To account for this, when computing  $\Delta \chi^2$ ,  $\Delta \text{AIC}$ , and  $\Delta \text{DIC}$ , we treat the Gaussian priors as effectively part of the likelihood. In practice what this means is that we evaluate them using the maximum posterior  $\mathcal{P}$  instead of the maximum likelihood  $\mathcal{L}$  and introduce a correction to account for differences in the flat prior contributions between models  $\mathcal{M}_X$  and  $\mathcal{M}_0$ .

To derive what the flat prior correction should be, we write the posterior  $\mathcal{P}$  for data  $\mathbf{d}$ , model  $\mathcal{M}$ , and parameters  $\Theta$  as

$$\mathcal{P}(\Theta|\mathbf{d}, \mathcal{M}) = \frac{\mathcal{L}(\mathbf{d}|\Theta, \mathcal{M}) \mathcal{G}(\Theta|\mathcal{M}) \pi^{\text{flat}}(\mathcal{M})}{\mathcal{Z}(\mathbf{d}|\mathcal{M})}, \quad (\text{F8})$$

where  $\mathcal{L}$  is the likelihood,  $\mathcal{G}$  is the product of the various of Gaussian priors on nuisance parameters, and  $\pi^{\text{flat}}$  is the  $\Theta$ -independent contribution to the prior from all flat priors. For the purpose of our model comparison statistics, we would like to define goodness-of-fit quantities like Eqs. (F4), (F5), and (F6), but depending on the product  $\mathcal{L}\mathcal{G}$  in place of just  $\mathcal{L}$ . The fact that  $\pi^{\text{flat}}$  does not depend on  $\Theta$  means that the best fit defined for  $\mathcal{L}\pi^{\text{flat}}$  will be the same vector  $\Theta_{\mathcal{M}}^{\text{MAP}}$  which maximizes  $\mathcal{P}$ . Suppressing arguments for conciseness, we note that a  $\chi^2$ -like posterior-based goodness-of-fit for model  $\mathcal{M}_X$  is

$$[\chi_{\mathcal{P}}^2]_X = -2 \ln (\mathcal{L}_X^{\text{MAP}} \mathcal{G}_X^{\text{MAP}}) - 2 \ln \pi_X^{\text{flat}} \quad (\text{F9})$$

$$\equiv [\chi_{\mathcal{L}\mathcal{G}}^2]_X - 2 \ln \pi_X^{\text{flat}}. \quad (\text{F10})$$

Model comparison between  $\mathcal{M}_X$  and  $\mathcal{M}_0$  therefore involves the comparison

$$\Delta \chi_{\mathcal{L}\mathcal{G}}^2 \equiv [\chi_{\mathcal{L}\mathcal{G}}^2]_X - [\chi_{\mathcal{L}\mathcal{G}}^2]_0 \quad (\text{F11})$$

$$= [\chi_{\mathcal{P}}^2]_X - [\chi_{\mathcal{P}}^2]_0 + 2 \ln \pi_X^{\text{flat}} - 2 \ln \pi_0^{\text{flat}}. \quad (\text{F12})$$

Using this expression, the  $\Delta \chi^2$  model comparison statistic reported in the main body of this paper is computed as:

$$\Delta \chi^2 \equiv -2 (\ln \mathcal{P}_X^{\text{MAP}} - \ln \mathcal{P}_0^{\text{MAP}}) \quad (\text{F13})$$

$$+ 2 (\ln \pi_X^{\text{flat}} - \ln \pi_0^{\text{flat}}). \quad (\text{F14})$$

Similar calculations for AIC and DIC result in the same flat-prior correction term.

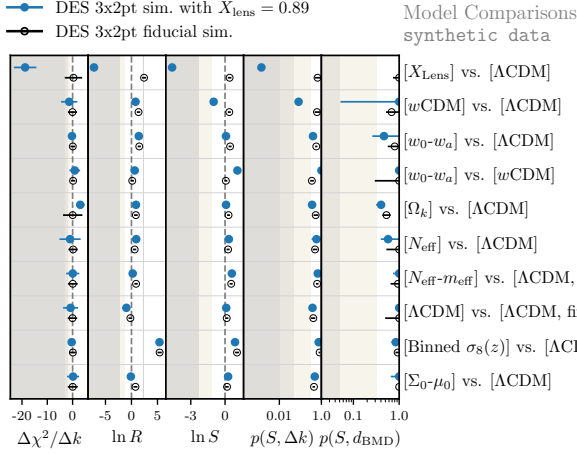


FIG. 24. Model comparison metrics when run on a synthetic data vector produced with  $X_{\text{Lens}} = 0.89$ . Points mark the median of 200 sampling variance estimates output by ANESTHETIC, and error bars (smaller than the points in most cases here) show the 16 and 84% quantiles.

### Appendix G: Testing beyond- $\Lambda$ CDM model response to $X_{\text{Lens}}$

We can use a synthetic data study similar to those in Sec. VB to assess how beyond- $\Lambda$ CDM models respond to a synthetic  $3 \times 2$ pt data vector produced with the redshift-independent parameter  $X_{\text{Lens}} \neq 1$ . As described in Sec. VC, the parameter  $X_{\text{Lens}}$  describes a mismatch between the galaxy bias detected by galaxy clustering versus that from galaxy-galaxy lensing. This is of interest because such a mismatch was found in the  $3 \times 2$ pt results for the alternative REDMAGIC lens sample studied in DES-Y3KP: when  $X_{\text{Lens}}$  is included in the REDMAGIC  $3 \times 2$ pt analysis, its preferred value is significantly less than the fiducial value of one. The effect is thought to be caused by residual systematics related to the lens sample selection, and remains a topic of active investigation. This indication of residual systematics ultimately motivated the choice of MagLim over REDMAGIC as the fiducial lens sample for DES-Y3KP, as this  $X_{\text{Lens}}$  effect is not present for the four-bin MagLim sample we are using for the analysis in this paper. Given this, we emphasize that this investigation is performed to understand the effects of such a systematic in extended model spaces, rather than as a validation test that our analysis must pass.

The test proceeds as follows: we generate a simulated data vector with  $X_{\text{Lens}} = 0.89$ , which approximates the preferred value from the REDMAGIC analysis. We then fit the  $X_{\text{Lens}}$ -contaminated synthetic data with our beyond- $\Lambda$ CDM models to see how they respond. Note that we use full MCMC chains for this as the FASTISMORE results indicated that the posterior shifts introduced by the  $X_{\text{Lens}}$  contamination were too large to be accurately captured through importance sampling.

Estimated model comparison metrics between  $\Lambda$ CDM and our extended models for such a contaminated data vector are shown in Fig. 24. Compared to the baseline simulated analysis, the  $X_{\text{Lens}}$  contamination causes model comparison met-

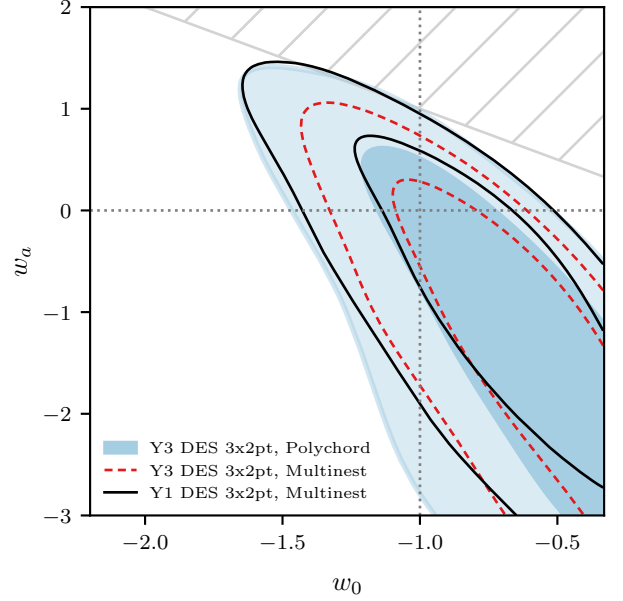


FIG. 25. Comparison between the DES Y3 and Y1  $3 \times 2$ pt constraints on dynamic dark energy parameters. Dashed red contours show that analyzing the Y3 data with Multinest sampler settings matching what those used in Y1 artificially narrow the Y3 posterior estimate.

rics to shift slightly in favor of extended models relative to  $\Lambda$ CDM, but none of these shifts are enough to cause an extended model to be strongly preferred to  $\Lambda$ CDM. Compare this to the  $\Delta\chi^2 \sim 18$  found when fitting a  $\Lambda$ CDM +  $X_{\text{Lens}}$  model (the correct model for the contamination). This is change in  $\chi^2$  is comparable to, if a bit smaller than, the  $\Delta\chi^2 \sim 25$  improvement found when fitting the REDMAGIC  $3 \times 2$ pt data in DES-Y3KP. The  $w$ CDM model shows perhaps the greatest sensitivity, and we do see  $> 1\sigma$  shifts upwards in the marginalized posterior on  $w$  when contaminated with  $X_{\text{Lens}}$ , which roughly agrees with the behavior seen in DES-Y3KP where the original REDMAGIC sample preferred  $w > -1$  (c.f. Fig. 12 of DES-Y3KP), in contrast to the fiducial MagLim sample. These findings are not unexpected. As was discussed in DES-Y3KP, the  $X_{\text{Lens}}$  effect seen for the initial Y3 REDMAGIC analysis appears to be much more consistent with a lens sample systematic than with new physics. The study presented here supports this, finding that the observed effect cannot be easily reproduced using the beyond- $\Lambda$ CDM models considered here.

### Appendix H: Comparison of Y1 and Y3 $w_0 - w_a$ constraints

Figure 25 shows a comparison between the DES Y3  $3 \times 2$ pt constraints on the  $w_0 - w_a$  dynamic dark energy parameters reported above in the body of the text to those previously published in DES-Y1Ext [50]. In that figure, the blue filled contour shows same result presented for DES Y3  $3 \times 2$ pt constraints in Fig. 4 of the main body of this paper, while the

black contours show the 68 and 95% confidence regions for the posterior estimate presented in DES-Y1Ext. We see that there is little change between our Y3  $w_0 - w_a$  results and those reported in the comparable DES Y1 analysis, although the modeling and analysis choices are similar in both cases.

One analysis choice that has an impact on this comparison is the sampler used for parameter estimation: we use Polychord in the present analysis while Multinest was used in the DES-Y1Ext analysis. To assess the impact of differences in this sampler choice on  $w_0 - w_a$  constraints we reanalyze the DES Y3  $3 \times 2$ pt data using the same Multinest sampler [203–205] settings used in DES-Y1Ext, but keeping all other analysis choices the same as in our fiducial Y3 analysis. One of the main motivations for switching to using the Polychord

sampler for DES Y3 analyses was that Multinest tends to undersample posterior tails [77], an effect which seems to be exacerbated in directions of parameter space where the posterior is more non-Gaussian. The impact of this undersampling is clearly visible for the  $3 \times 2$ pt  $w_0 - w_a$  constraints in Fig. 25, with the dashed red Multinest contours suggesting tighter constraints than the (more correct) blue Polychord contours and than the DES-Y1Ext Multinest constraints. Quantitatively, Multinest underestimates the width of the DES Y3  $3 \times 2$ pt marginalized 68 and 96% confidence intervals for  $w_p$  by 7 and 21%, respectively, compared to the same quantities estimated using Polychord. This suggests that the Y1  $3 \times 2$ pt  $w_0 - w_a$  constraints may be at least somewhat artificially tightened due to the use of Multinest.