

(Lack of) Cosmological evidence for dark radiation after Planck

Licia Verde,^{a,b} Stephen M. Feeney,^c Daniel J. Mortlock^{d,e} and Hiranya V. Peiris^{c,f}

^aICREA & ICC, Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franques 1, Barcelona 08028, Spain

^bTheory Group, Physics Department, CERN, CH-1211, Geneva 23, Switzerland

^cDepartment of Physics and Astronomy, University College London, London WC1E 6BT, U.K.

^dAstrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.

^eDepartment of Mathematics, Imperial College London, London SW7 2AZ, U.K.

^fKavli Institute for Theoretical Physics, Kohn Hall, University of California, Santa Barbara, CA 93106, USA

E-mail: liciaverde@icc.ub.edu, stephen.feeney.09@ucl.ac.uk, mortlock@ic.ac.uk, h.peiris@ucl.ac.uk

Abstract. We use Bayesian model comparison to determine whether extensions to Standard-Model neutrino physics – primarily additional effective numbers of neutrinos and/or massive neutrinos – are merited by the latest cosmological data. Given the significant advances in cosmic microwave background (CMB) observations represented by the *Planck* data, we examine whether *Planck* temperature and CMB lensing data, in combination with lower redshift data, have strengthened (or weakened) the previous findings. We conclude that the state-of-the-art cosmological data do not show evidence for deviations from the standard (Λ CDM) cosmological model (which has three massless neutrino families). This does not mean that the model is necessarily correct – in fact we know it is incomplete as neutrinos are not massless – but it does imply that deviations from the standard model (e.g., non-zero neutrino mass) are too small compared to the current experimental uncertainties to be inferred from cosmological data alone.

Contents

1	Introduction	1
2	Methods and datasets	2
2.1	Bayes factor	2
2.2	Savage-Dickey Density Ratio	3
2.3	Profile likelihood ratio	5
2.4	Data	5
3	Results	6
3.1	Evidence	7
3.2	Profile likelihood ratios and comparison with Evidence	9
3.3	Exploring degeneracies with other interesting parameters: the case of n_s	10
4	Conclusions and future avenues	11

1 Introduction

Exploring non-standard neutrino properties (including a significantly non-zero mass) requires extending both the concordance cosmological model, Λ CDM, and the Standard Model of particle physics, and is therefore an issue of model comparison rather than parameter estimation. Model comparison must be performed within the Bayesian framework to be self-consistent [1], and the relevant quantity to consider in this context is the model-averaged likelihood, referred to here as the Bayesian *Evidence*, E . Under the assumption of equal *a priori* model probabilities, the ratio of Evidence¹ values for two models, given the same data, quantifies the relative odds of these models being the correct description of the observations. The use of the Bayesian model comparison framework in cosmology has become popular over the last two decades: e.g., see Refs. [2–7] and references therein. In previous work [8], we considered the cosmological Bayesian Evidence for non-standard neutrino properties using pre-*Planck* cosmological data. In this paper we update our previous work by considering a compilation of post-*Planck* [9] cosmological data, which have significantly extended the state-of-the-art compared with the data used in our previous analysis.

The Standard Model of particle physics has three massless neutrinos. Beyond-Standard-Model physics (or uncomfortable fine-tuning) is needed to give neutrinos a non-zero mass, and extensions of the Standard Model include the possibility of more than three neutrino species. The standard cosmological model, Λ CDM, also has three massless neutrino families, but quantifies their effects through the effective number of species, $N_{\text{eff}} = 3.046$. This differs from the number of neutrino species, $N_\nu = 3$, to account for QED effects, for neutrinos being not completely decoupled during electron-positron annihilation and other small effects (see, e.g., Refs. [10, 11]). For cosmological observations, a deviation from the Standard Model prediction for N_{eff} does not necessarily imply new neutrino physics: e.g., any non-standard early-Universe expansion history due to non-standard energy-density can be parametrized in terms of N_{eff} . This “dark radiation” has been the subject of great interest in the pre-*Planck* years: for a review, see Ref. [12] and references therein, and the discussion and references in Ref. [8].

The pre-*Planck* cosmological data have been often interpreted as supporting the case for dark radiation, as well as one or more sterile neutrinos. Post-*Planck*, the data leave less freedom to the sterile-neutrino interpretation [13], but Bayesian parameter estimates of models with N_{eff} as a free parameter still leave room for dark radiation (see, e.g., Refs. [14–20]). The official *Planck* analysis (and other work considering *Planck* data) concentrated on parameter estimation and did not address the model comparison issue. It is therefore important to consider the newly-available data in the

¹We capitalize the Bayesian Evidence to distinguish it from the colloquial “evidence”.

framework of model comparison, and examine how these findings might depend on the assumed priors. In addition we note that, while in the standard cosmological model the neutrinos have traditionally been assumed to be massless, the *Planck* collaboration chose their baseline, standard Λ CDM model to have a non-zero total neutrino mass: $M_\nu = 0.06$ eV [13], which is close to the minimum value allowed by neutrino-oscillation experiments (see, e.g., Ref. [21] and references therein). Because of the significant increase in the precision of recent cosmological data, this choice implies small shifts in the best-fit cosmological parameters compared to a model in which $M_\nu = 0$. It is interesting to consider whether current cosmological data exhibit any evidence for this “paradigm shift” (though this choice is well-motivated by non-cosmological data).

This paper is organized as follows. In Sec. 2 we describe the methods used in our analysis, their interpretation, and the datasets we consider. In Sec. 3 we present our results, and in Sec. 4 we present our conclusions and discuss future avenues to further constrain neutrino properties in the context of cosmological model comparison.

2 Methods and datasets

We wish to answer the following question: do the latest cosmological data require the inclusion of extra parameters (describing non-standard neutrino properties) beyond the simple Λ CDM model? Rather than being an issue of parameter estimation – i.e., the determination of the most probable values for the extra parameter(s) within the context of a single model – this is a question of model comparison.

Within a Bayesian framework the key model comparison quantity is the Bayes factor, which is the ratio of the Evidence values for two different models (see Sec. 2.1 below). In general, the Evidence is the result of a multi-dimensional integral over the model parameters, θ , the evaluation of which can be computationally expensive, and specialized algorithms have been developed to compute it efficiently [22, 23]. Here, however, the Bayes factors can be computed much more efficiently because the models we are comparing have some parameters in common, and indeed are nested (see Sec. 2.2 below). The parameter space $\theta = (\phi, \psi)$ is partitioned into the common parameters ϕ (in this case, those of vanilla Λ CDM: the physical densities of baryons and cold dark matter, $\Omega_b h^2$ and $\Omega_c h^2$; the angular scale of the sound horizon at last-scattering, Θ_A ; the integrated optical depth to last-scattering, τ ; and the matter power spectrum spectral slope and amplitude, n_s and A_s) and the extra parameters ψ (in this case, those describing the additional neutrino properties, N_{eff} , M_ν , etc.). The question asked above is addressed by determining whether the data support a model, M_2 , in which the extra parameters are not restricted to their fiducial Λ CDM values.

As with all Bayesian methods, the Bayes factor between these two models depends on the prior on the model parameters. While in some cases the priors can be physically motivated, this is not always possible; in the latter case, the Bayesian answer to the model comparison question calculated with specific prior choices cannot be considered definitive. We will therefore also consider a statistic – the profile likelihood ratio (PLR) – which depends only on the likelihood, and thus is prior-independent. While the PLR does not provide a self-consistent model-selection paradigm, it is useful for assessing whether the Bayesian results are driven by the prior choice or the data.

2.1 Bayes factor

The probability distribution (Pr) for a set of parameters, θ , given a model, M , and data, \mathbf{d} , is the posterior, $P = \text{Pr}(\theta|\mathbf{d}, M)$. Bayes’ theorem relates the posterior to the likelihood, $\mathcal{L} \equiv \text{Pr}(\mathbf{d}|\theta, M)$, via the prior, $\pi \equiv \text{Pr}(\theta|M)$:

$$\text{Pr}(\theta|\mathbf{d}, M) = \frac{\text{Pr}(\theta|M) \text{Pr}(\mathbf{d}|\theta, M)}{\text{Pr}(\mathbf{d}|M)}. \quad (2.1)$$

The Bayesian Evidence or model-averaged likelihood, $E \equiv \text{Pr}(\mathbf{d}|M)$, normalizes the parameter posterior, and is given by

$$E = \int \text{Pr}(\theta|M) \text{Pr}(\mathbf{d}|\theta, M) d\theta. \quad (2.2)$$

$ \ln(E_1/E_2) $	interpretation	betting odds
< 1	not worth a bare mention	$< 3 : 1$
$1 - 2.5$	substantial	$\sim 3 : 1$
$2.5 - 5$	strong	$> 12 : 1$
> 5	highly significant	$> 150 : 1$

Table 1. The slightly modified Jeffreys’ scale we use for interpreting the Evidence ratios.

Using this expression we can write the ratio of probabilities of two models in the light of the data as

$$\frac{\Pr(M_1|\mathbf{d})}{\Pr(M_2|\mathbf{d})} = \frac{\Pr(M_1) \Pr(\mathbf{d}|M_1)}{\Pr(M_2) \Pr(\mathbf{d}|M_2)} = \frac{\Pr(M_1) E_1}{\Pr(M_2) E_2}, \quad (2.3)$$

where the *a priori* probability ratio of the two models, $\Pr(M_1)/\Pr(M_2)$, is typically set to unity, and $\Pr(\mathbf{d}|M_1)/\Pr(\mathbf{d}|M_2) = E_1/E_2$ is the Bayes factor between the two models. The Bayes factor, or ratio of the Evidence values for two models given the same data, therefore expresses the relative odds that these models are responsible for the observed state of the Universe.

It is important to note here that Bayesian model comparison does not automatically reward *simpler* models, but rather rewards *predictive* models (see, e.g., Ref. [24]). If an extension to Λ CDM predicts observables (e.g., CMB power spectra) that are indistinguishable from those of Λ CDM (and the two models’ likelihoods are hence equal), then the Evidence values for the two models are equal. More generally, consider a simple model, M_1 , and an extended model, M_2 . If the extra parameters in the extended model have no influence on the likelihood, so that $\Pr(\mathbf{d}|\phi, \psi, M_2) = \Pr(\mathbf{d}|\phi, M_1)$, and the marginalized prior for the common parameters in the extended model, $\Pr(\phi|M_2) = \int \Pr(\phi, \psi|M_2) d\psi$, is the same as their prior in the simpler model, $\Pr(\phi|M_1)$, then the Evidence for the extended model is given by

$$\begin{aligned} E_2 &= \int \int \Pr(\phi, \psi|M_2) \Pr(\mathbf{d}|\phi, \psi, M_2) d\phi d\psi \\ &= \int \left[\int \Pr(\phi, \psi|M_2) d\psi \right] \Pr(\mathbf{d}|\phi, M_1) d\phi \\ &= \int \Pr(\phi|M_1) \Pr(\mathbf{d}|\phi, M_1) d\phi \\ &= E_1. \end{aligned}$$

Within the context of Bayesian model comparison a model does not have an intrinsic complexity; it is only the ability of a model to predict the results of measurements that is important. If the available observations are equally likely within the context of two models then neither is preferred by the data, regardless of their internal structure.

In situations where the evidence values differ we use a slightly modified Jeffreys’ scale [25, 26] to interpret our Evidence ratios. This scale, which classifies Evidence ratios from “not worth a bare mention” to “highly significant”, is defined in Table 1). Care should be taken when using thresholded scales such as the Jeffreys’ scale, which introduce sharp decision-making boundaries. It is used here only as a loose classification of the strength of the Evidence ratios, as we are comparing a large number of model and dataset combinations.

2.2 Savage-Dickey Density Ratio

The extensions to Λ CDM considered in this work are clearly related to Λ CDM itself: Λ CDM is a simpler version of the extended models, and is in fact “nested” within the extensions. A simpler model M_1 can be considered to be nested within an extended model M_2 if:

- the parameters which describe model M_2 are separable into two disjoint subsets: the base parameters, ϕ , fully describing M_1 , and some extra parameters, ψ , unique to M_2 ;

- there exists a combination of the extra parameters, $\boldsymbol{\psi} = \boldsymbol{\psi}_1$, for which the second model reduces identically to the simpler model so that the likelihood $\Pr(\mathbf{d}|\boldsymbol{\phi}, M_1) = \Pr(\mathbf{d}|\boldsymbol{\phi}, \boldsymbol{\psi}_1, M_2)$ for all \mathbf{d} ;
- the prior distributions of the extra parameters in M_2 are separable from the prior distributions of the base parameters: $\Pr(\boldsymbol{\phi}, \boldsymbol{\psi}|M_2) = \Pr(\boldsymbol{\phi}|M_2)\Pr(\boldsymbol{\psi}|M_2)$, with $\Pr(\boldsymbol{\psi}_1|M_2) > 0$;
- the prior distributions of the common parameters are the same: $\Pr(\boldsymbol{\phi}|M_1) = \Pr(\boldsymbol{\phi}|M_2)$.

In the case of nested models, the Bayes factor is

$$\frac{E_1}{E_2} = \frac{\Pr(\mathbf{d}|M_1)}{\Pr(\mathbf{d}|M_2)} = \frac{\int \Pr(\boldsymbol{\phi}'|M_1) \Pr(\mathbf{d}|\boldsymbol{\phi}', M_1) d\boldsymbol{\phi}'}{\int \Pr(\boldsymbol{\phi}'', \boldsymbol{\psi}''|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}'', \boldsymbol{\psi}'', M_2) d\boldsymbol{\phi}'' d\boldsymbol{\psi}''} \quad (2.4)$$

$$= \frac{\int \Pr(\boldsymbol{\phi}'|M_1) \Pr(\mathbf{d}|\boldsymbol{\phi}', \boldsymbol{\psi}_1, M_2) d\boldsymbol{\phi}'}{\int \Pr(\boldsymbol{\phi}''|M_1)\Pr(\boldsymbol{\psi}''|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}'', \boldsymbol{\psi}'', M_2) d\boldsymbol{\phi}'' d\boldsymbol{\psi}''}, \quad (2.5)$$

where in the second line we have used the fact that the models are nested, and care has been taken to avoid confusing the distinct dummy integration variables in the numerator and the denominator. Multiplying both numerator and denominator by $\Pr(\boldsymbol{\psi}_1|M_2)$ then gives

$$\frac{E_1}{E_2} = \frac{1}{\Pr(\boldsymbol{\psi}_1|M_2)} \frac{\int \Pr(\boldsymbol{\phi}'|M_1) \Pr(\boldsymbol{\psi}_1|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}', \boldsymbol{\psi}_1, M_2) d\boldsymbol{\phi}'}{\int \Pr(\boldsymbol{\phi}''|M_1) \Pr(\boldsymbol{\psi}''|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}'', \boldsymbol{\psi}'', M_2) d\boldsymbol{\phi}'' d\boldsymbol{\psi}''}. \quad (2.6)$$

By comparison, the normalized parameter posterior for $(\boldsymbol{\phi}, \boldsymbol{\psi})$ under the second model is

$$\Pr(\boldsymbol{\phi}, \boldsymbol{\psi}|\mathbf{d}, M_2) = \frac{\Pr(\boldsymbol{\phi}|M_1) \Pr(\boldsymbol{\psi}|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}, \boldsymbol{\psi}_1, M_2)}{\int \Pr(\boldsymbol{\phi}''|M_1) \Pr(\boldsymbol{\psi}''|M_2) \Pr(\mathbf{d}|\boldsymbol{\phi}'', \boldsymbol{\psi}'', M_2) d\boldsymbol{\phi}'' d\boldsymbol{\psi}''} \quad (2.7)$$

for all values of $\boldsymbol{\psi}$ (and, in particular, for $\boldsymbol{\psi} = \boldsymbol{\psi}_1$). Hence

$$\frac{E_1}{E_2} = \frac{\int \Pr(\boldsymbol{\phi}', \boldsymbol{\psi}_1|\mathbf{d}, M_2) d\boldsymbol{\phi}'}{\Pr(\boldsymbol{\psi}_1|M_2)}. \quad (2.8)$$

But integrating a posterior distribution over some sub-set of the parameters is simply marginalization, and in this case the integral in the numerator reduces to the marginalized posterior $\Pr(\boldsymbol{\psi}_1|\mathbf{d}, M_2)$. Hence the evidence ratio between two nested models is given by

$$\frac{E_1}{E_2} = \frac{\Pr(\boldsymbol{\psi}_1|\mathbf{d}, M_2)}{\Pr(\boldsymbol{\psi}_1|M_2)} = \frac{\Pr(\boldsymbol{\psi}|\mathbf{d}, M_2)}{\Pr(\boldsymbol{\psi}|M_2)} \Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}_1}. \quad (2.9)$$

This is the Savage–Dickey density ratio (SDDR; Ref. [27]²). Using the SDDR makes it possible to perform rigorous model comparison between nested models without the need for evaluating numerically-intensive multi-dimensional integrals over the two models' parameters.

It may seem surprising that such an apparently complicated ratio of integrals has such a simple expression, but the interpretation of this final result is clear. The simpler nested model is preferred if, within the context of the more-complicated model, the relevant data result in an increased probability that $\boldsymbol{\psi} = \boldsymbol{\psi}_1$; the more complicated model is preferred if the data imply that $\boldsymbol{\psi} = \boldsymbol{\psi}_1$ is disfavoured compared with other values for the extra parameters in the second model. Note also that this gives an alternative derivation of the result given above for the case in which the extra parameters have no influence on the likelihood; in that case the distributions $\Pr(\boldsymbol{\psi}|\mathbf{d}, M_2)$ and $\Pr(\boldsymbol{\psi}|M_2)$ must be the same, again yielding $E_2 = E_1$.

Critically, the SDDR given in Eq. 2.9 can be estimated from samples drawn from the posterior distribution $\Pr(\boldsymbol{\theta}|\mathbf{d}, M_2)$, such as a Markov Chain Monte Carlo (MCMC) outputs. For detailed and high-precision implementations see, e.g., Refs. [31, 32]. We will use a simpler implementation which still yields very accurate results in the relevant setting, as discussed in Sec. 3.1.

²Dickey attributed this result to Savage, hence the name. The implications of this result are further explored in e.g., Refs. [7, 28–30].

2.3 Profile likelihood ratio

In the absence of well-motivated parameter priors, we can use the likelihood alone to assess the significance of any deviations of the extra parameters (ψ) from their fiducial values (ψ_1). This can be seen by considering an extreme alternative model in which the extra parameters are fixed at some value ψ_2 , implying a parameter prior of the form $\Pr(\phi, \psi|M_2) = \Pr(\phi|M_2) \delta_D(\psi - \psi_2)$. The Evidence for this model immediately simplifies to $E_2 = \int \Pr(\phi'|M_2) \Pr(\mathbf{d}|\phi', \psi_2) d\phi'$, and would be maximized if $\Pr(\phi|M_2) = \delta_D[\phi - \phi_{\text{ML}}(\psi_2, \mathbf{d})]$, where $\phi_{\text{ML}}(\psi_2, \mathbf{d})$ is the maximum likelihood (ML) value of ϕ under the restriction that $\psi = \psi_2$. While it would be unreasonable to assess a model by tuning it to the same data being used to test it, this does provide an upper bound on the Evidence, as $E_2 \leq \Pr[\mathbf{d}|\phi_{\text{ML}}(\psi_2, \mathbf{d}), \psi_2] = \mathcal{L}_{\text{max}}(\psi_2)$.

It is now possible to examine how $\mathcal{L}_{\text{max}}(\psi_2)$ depends on the value of ψ_2 . The maximum possible value of $\mathcal{L}_{\text{max}}(\psi_2)$ occurs if $\psi_2 = \psi_{\text{ML}}(\mathbf{d})$, the overall maximum-likelihood model, in which case $\mathcal{L}_{\text{max}}(\psi_2) = \Pr[\mathbf{d}|\phi_{\text{ML}}(\mathbf{d}), \psi_{\text{ML}}(\mathbf{d})] = \mathcal{L}_{\text{max}}$. The normalized ratio

$$\frac{\Pr[\mathbf{d}|\phi_{\text{ML}}(\psi_2, \mathbf{d}), \psi_2]}{\Pr[\mathbf{d}|\phi_{\text{ML}}(\mathbf{d}), \psi_{\text{ML}}(\mathbf{d})]} = \frac{\mathcal{L}_{\text{max}}(\psi_2)}{\mathcal{L}_{\text{max}}} \quad (2.10)$$

then gives a heuristic assessment of how plausible a model with $\psi = \psi_2$ is given the data. This is known as the profile likelihood ratio (PLR; [33]).

The PLR has an interpretation similar to the $\Delta\chi^2$ for nested models, where the effective chi-square is identified with $-2 \ln \Pr(\mathbf{d}|\phi, \psi, M_2)$ and, by construction, is prior-independent. Assuming Gaussian statistics one may associate ratios of 0.5 and 2.0 to one- and two- σ confidence regions. If, e.g., the PLR at a non-standard value of N_{eff} were two greater than the PLR at $N_{\text{eff}} = 3.046$, one could claim $\sim 2\sigma$ evidence in favor of the $\Lambda\text{CDM} + N_{\text{eff}}$ model. It must be noted that these confidence intervals may not have strict frequentist coverage, especially if the likelihood is far from Gaussian; nevertheless, this quantity allows us to assess whether the Evidence results are driven by the data or by the prior.

In practice, the PLR is computed from the output of Monte Carlo Markov chains with the approximation that the conditional maximum likelihood is computed in bins for the interesting parameter (see, e.g., Ref. [8, 34, 35].) We verify that the dependence on the bin size is small.

2.4 Data

The *Planck* satellite mission [9] recently released its temperature data from 2.6 surveys of the sky (15 months). Along with the data, the *Planck* Collaboration also released the MCMCs used to sample from the space of possible cosmological parameters and generate estimates of the posterior mean of each parameter of interest, along with a confidence interval. Here, we use the publicly-available outputs of the *Planck* Collaboration's MCMCs.

Several combinations of the *Planck* data were analyzed by the Collaboration, and a large number of MCMCs are therefore available. The minimum dataset uses only *Planck* temperature data (*Planck-lowl*, here shortened to *Planck*) in the multipole range $\ell < 2500$. Since there are no *Planck* polarization data in this first cosmological data release to constrain the optical depth to reionization, τ , WMAP polarization data [36] (*lowLike* or *WP*) at low ℓ are also included; in some cases a τ prior (*tauprior*) is used instead, but we do not use it here. The *Planck* data analysis has also produced a reconstruction of the CMB lensing potential through the measurement of the temperature-anisotropy four-point function. The lensing potential probes the amplitude of the large-scale cosmological structure integrated all the way to the present time, thus helping break cosmological degeneracies that are intrinsic to the primary CMB anisotropies. When the information from the lensing potential is included it is referred to as *lensing*. Note that the effects of lensing on the primary anisotropies (smoothing of the higher-order acoustic peaks of the temperature power spectrum and the change in the damping tail) are always included.

Ground-based telescopes (the Atacama Cosmology Telescope, ACT [37], and the South Pole Telescope, SPT [38]) have mapped the sky over small areas at CMB frequencies. These experiments have mapped the CMB damping tail but, more importantly in the context of the *Planck* data analysis,

data	shorthand	reference
<i>Planck</i> temperature	Planck	[39]
WMAP low ℓ polarization	lowLike or WP	[36]
Prior on optical depth	tauprior	[39, 40]
<i>Planck</i> reconstructed lensing potential	lensing	[41]
ACT and SPT	highL	[37, 38]
Baryon acoustic oscillations compilation	BAO	[42–46]
Hubble constant	HST	[47]
Supernovae	SNLS	[48]
Supernovae	Union2	[49]

Table 2. Summary of the combinations of datasets, their abbreviation and relevant references.

have mapped the foregrounds with higher resolution and lower noise than *Planck* (albeit not over the full sky). These data (named *highL*) can therefore be used with *Planck* data to better constrain the foreground-model parameters and thus reduce degeneracies these parameters may have with the cosmology. Other lower redshift observations can also be used to break cosmological degeneracies. The full set of such data sources used in the *Planck* analysis is summarized in Table 2.

In this paper our “Gold” set will be *Planck* temperature data combined with WMAP polarization information at low ℓ . We will also report results for other combinations (where available), but we prefer to highlight the Gold dataset for the following reasons. Firstly, it involves only high-quality CMB observations whose interpretation is based on well-understood, linear physics. Secondly, the combination of *Planck* temperature data and WMAP polarization data (Planck+WP) with *highL* data predicts a value for the lensing amplitude which is about 2σ higher than the value measured from the convergence power spectrum [13]. As the origin of this tension is not yet fully understood, we prefer not to include *lensing* and *highL* in our Gold dataset. There has also been significant discussion on the fact that the value of the Hubble constant extrapolated from *Planck* is in tension with the local measurement [13]. We prefer not to combine datasets that are slightly in tension in our Gold set.

Other datasets considered by the *Planck* Collaboration in combination with CMB data include measurements of the baryon acoustic oscillation feature (BAO) [42–46] and two sets of supernovae data, SNLS [48] and Union2 [49]. While BAO data are in good agreement with *Planck*, the two supernovae datasets give slightly different results when combined with CMB data. For this reason, the two supernovae datasets are always considered separately.

Although we focus on the Gold set, especially in the discussion of the results and in figures, we report full results for a range of data combinations beyond the Gold set.

3 Results

We consider the following extensions to the Λ CDM model:

1. the addition of a non-standard effective number of neutrino species, Λ CDM+ N_{eff} (within which Λ CDM is nested at $N_{\text{eff}} = 3.046$);
2. the addition of a non-standard neutrino number and a primordial Helium abundance (Y_P), not fixed by Big Bang nucleosynthesis, Λ CDM+ $N_{\text{eff}} + Y_P$ (within which Λ CDM is nested at $N_{\text{eff}} = 3.046$, $Y_P = 0.24$);
3. a model with three massive neutrinos, Λ CDM+ M_ν , where M_ν is the sum of the individual neutrino masses (within which Λ CDM is nested at $M_\nu = 0$, or 0.06 eV following Ref. [13]);
4. the addition of both neutrino mass and a non-standard effective number of neutrinos, Λ CDM+ $M_\nu + N_{\text{eff}}$ (within which Λ CDM is nested at $N_{\text{eff}} = 3.046$, $M_\nu = 0$ or 0.06); and

parameter	ϕ						ψ			
	$\Omega_b h^2$	$\Omega_c h^2$	Θ_A	τ	n_s	$\ln(10^{10} A_s)$	M_ν	N_{eff}	Y_P	$m_{\nu_{\text{sterile}}}^{\text{eff}}$
prior minimum	0.005	0.001	0.005	0.01	0.9	2.7	0	0.05	0.1	0
prior maximum	0.1	0.99	0.1	0.8	1.1	4	5	10	0.5	3

Table 3. Prior ranges used for the cosmologically relevant parameters. The priors employed are those of Ref. [13]; see their Table 1 for more details.

5. a model with one massive sterile neutrino in addition to the three active neutrinos (two massless and one massive), $\Lambda\text{CDM}+N_{\text{eff}}+m_{\nu_{\text{sterile}}}^{\text{eff}}$ (within which ΛCDM is nested at $N_{\text{eff}} = 3.046$, $m_{\nu_{\text{sterile}}}^{\text{eff}} = 0$) – the two parameters describe a sterile neutrino with freely-varying mass and temperature.

The priors used are those of Ref. [13], which are reported in their Table 1; we summarize the priors on the cosmologically-interesting parameters here in Table 3.

3.1 Evidence

As motivated in Sec. 2.2 we estimate the Evidence with the SDDR. The numerical implementation of this estimate, when evaluated from MCMCs, introduces two approximations: (1) the MCMC is a finite sampling of the posterior, and (2) the conditional maximum of the posterior is calculated for binned values of the parameter of interest. We quantify the error on the Evidence ratio introduced by each of these effects. The MCMC sampling effect can be estimated in a jackknife-like fashion by evaluating the scatter on the evidence computed from sub-sections of the MCMCs. The binning effect is evaluated by changing the bin size. We find that both these effects contribute errors of $\sim \pm 0.02$ to the log-evidence ratio for most cases. Errors can reach ± 0.1 for thinned, post-processed (importance-sampled) chains, but even these errors are small compared to the relevant $\ln(E_1/E_2)$.

The MCMCs provide very precise estimates of the Evidence ratio, despite not being designed with Evidence calculation in mind. This is a consequence of our studying nested models, and specifically models nested at parameter values fairly close to the peak in the posterior. In these circumstances, the MCMCs characterize the posterior at the nested value very well, providing an accurate and precise estimate of the SDDR. Should the nested value be far from the posterior peak we would expect both the accuracy and precision of the SDDR to suffer, although arguably in this case the model comparison results would be definitively in favor of the more-complex model.

Further confidence in the performance of the SDDR applied to MCMCs can be obtained by direct comparison to the Evidence as calculated by nested sampling. In previous work [8], we used MultiNest [50, 51] to calculate the Evidence for various models with extended neutrino physics in the light of CMB observations by WMAP [52] and SPT [53]. Both of these collaborations have performed parameter-estimation for the same (or similar³) models, and have released the MCMCs from which their parameter constraints are obtained. Calculating the SDDR using these chains, we find that the typical difference between the log-Evidence ratios ($\ln[E_{\Lambda\text{CDM}+}/E_{\Lambda\text{CDM}+N_{\text{eff}}}]$) obtained via the SDDR and nested sampling is 0.1–0.3, in line with the estimated error in the MultiNest calculation of ~ 0.3 .

The complete set of Evidence ratios comparing the base ΛCDM model with the more-complex models is reported in Table 4. Fig. 1 also shows a visual comparison of a selection of the table entries. Concentrating first on the $\Lambda\text{CDM}+N_{\text{eff}}$ model comparisons, we see that the Evidence in favor of the simplest model (ΛCDM) is “substantial” to “strong”, except when the H_0 measurement is added. The high local measurements of H_0 and the low extrapolations from *Planck* are known to be discrepant at the $\sim 2\sigma$ level, a tension that can be relieved by increasing the effective number of neutrino species (and hence the expansion rate at both early and late times [54]). Nevertheless, even when considering datasets known to be in tension, there is still no substantial Evidence in favor of the more complex $\Lambda\text{CDM}+N_{\text{eff}}$ model. As anticipated in Ref. [8], the CMB lensing signal offers a

³The prior on N_{eff} employed by WMAP extends to slightly lower values than that used by SPT and the previous work; however, the likelihood is very low in this region of parameter space ($0 < N_{\text{eff}} < 1$), and this region can therefore be excluded from the SDDR calculation without significantly changing the results.

		evidence ratio: $\ln(E_{\Lambda\text{CDM}}/E_{\text{extension}})$					
model extension	data	base	+BAO	+ H_0	+SNLS	+Union2	+lensing
$\Lambda\text{CDM}+M_\nu$ (vs $M_\nu = 0$ eV)	Planck	2.12	3.25 P	3.73 P	3.42 P	2.97 P	0.38 P
	Planck+WP	2.65	3.56 P	3.88 P	3.62 P	3.25 P	
	+ lensing	1.88	3.27 P	3.83 P	3.48 P		
	Planck+WP+highL	3.00	3.63 P	3.93 P	3.70 P	3.39 P	
$\Lambda\text{CDM}+M_\nu$ (vs $M_\nu = 0.06$ eV)	Planck	2.01	3.17 P	3.40 P	3.20 P	2.85 P	0.33 P
	Planck+WP	2.53	3.35 P	3.46 P	3.32 P	3.06 P	
	+ lensing	1.83	3.24 P	3.39 P	3.20 P		
	Planck+WP+highL	2.90	3.43 P	3.51 P	3.40 P	3.22 P	
$\Lambda\text{CDM}+N_{\text{eff}}$	Planck	-0.43	1.78 P	-1.44 P	-1.89 P	-0.21 P	1.27 P
	Planck+WP	1.62	1.90	-0.81	1.49	0.96	1.93 P
	+ lensing P	1.93	2.36	-0.02	1.91	1.60	
	Planck+WP+highL	2.00	2.28	0.14 P	1.52 P	1.94 P	2.17
$\Lambda\text{CDM}+N_{\text{eff}}+Y_p$	Planck+WP	1.76	2.09 P	1.10 P			1.72 P
	Planck+WP+highL	1.75	2.11 P	0.48 P			1.78 P
$\Lambda\text{CDM}+N_{\text{eff}}+M_\nu$	Planck+WP	4.36	5.11 P	2.80 P			3.87 P
	Planck+WP+highL	5.05	5.59 P	3.73 P			4.68 P
$\Lambda\text{CDM}+N_{\text{eff}}+m_{\nu\text{sterile}}^{\text{eff}}$	Planck+WP+highL	3.52	3.72 P	2.29 P			3.38 P

Table 4. Evidence results for all of the model-dataset combinations we consider. A “P” next to the number or dataset indicates that the results have been obtained by importance sampling (i.e., post-processing) the MCMCs. Note that the extra datasets (BAO, supernovae, etc.) are always added one-by-one to the base combination. Positive numbers mean that the simpler model is preferred, negative numbers that the more-complicated model is preferred. The errors on the evidence ratios are in most cases about ± 0.02 , reaching ± 0.1 for thinned, post-processed chains.

powerful constraint for the simple N_{eff} extension, having as much statistical power as the combined BAO dataset. By comparing our Fig. 1 with Fig. 6 of Ref. [8] we can appreciate at a glance that, while with pre-*Planck* data the Evidence was inconclusive, it is now substantially in favor of ΛCDM , except when the discrepant H_0 is included.

Also particularly interesting are the Evidence ratios for adding massive neutrinos to ΛCDM , as observations of neutrino oscillations imply neutrinos are not massless. The question that the Evidence ratios address is therefore whether the masses are cosmologically relevant, and the answer from current cosmological data is a resounding “no”, stronger even than the results from WMAP [8]. This can be seen by comparing our Fig. 1 with its pre-*Planck* equivalent: Fig. 6 in Ref. [8]. For all pre-*Planck* dataset combinations the Evidence was indecisive, only becoming substantially in favor of ΛCDM when the H_0 measurement and the CMB lensing signal from SPT were included. Now the Evidence is strongly in favor of ΛCDM for all combinations in which an extra dataset is added to the CMB.

Let us note again that Bayesian model comparison rewards *predictive* rather than *simple* models. If $\Lambda\text{CDM}+M_\nu$ predicted power spectra that fit the data as well as those predicted by ΛCDM then the Evidence values for the two models would be equal. This is not the case here: $\Lambda\text{CDM}+M_\nu$ is disfavored because, on average, summed neutrino masses in the specified prior range provide a worse fit to the data than ΛCDM . Cosmological data do not yet have sufficient precision such that the likelihood for $M_\nu > 0$ is significantly and sufficiently better than that for $M_\nu = 0$ to tilt the verdict of the Evidence ratio.

It is interesting to note that adding massive neutrinos to ΛCDM is particularly disfavored by the Evidence for the data combination(s) that involve H_0 . This is again due to the mild tension between the CMB-extrapolated H_0 value and the local measurement of the same quantity, which is worsened by increasing the neutrino mass.

When even more complex models are considered, with two extra parameters rather than one (i.e., adding N_{eff} and M_ν , or N_{eff} and $m_{\nu\text{sterile}}^{\text{eff}}$) the preference of the Evidence ratio for the standard ΛCDM model increases. This is another consequence of the fact that the standard ΛCDM model is a

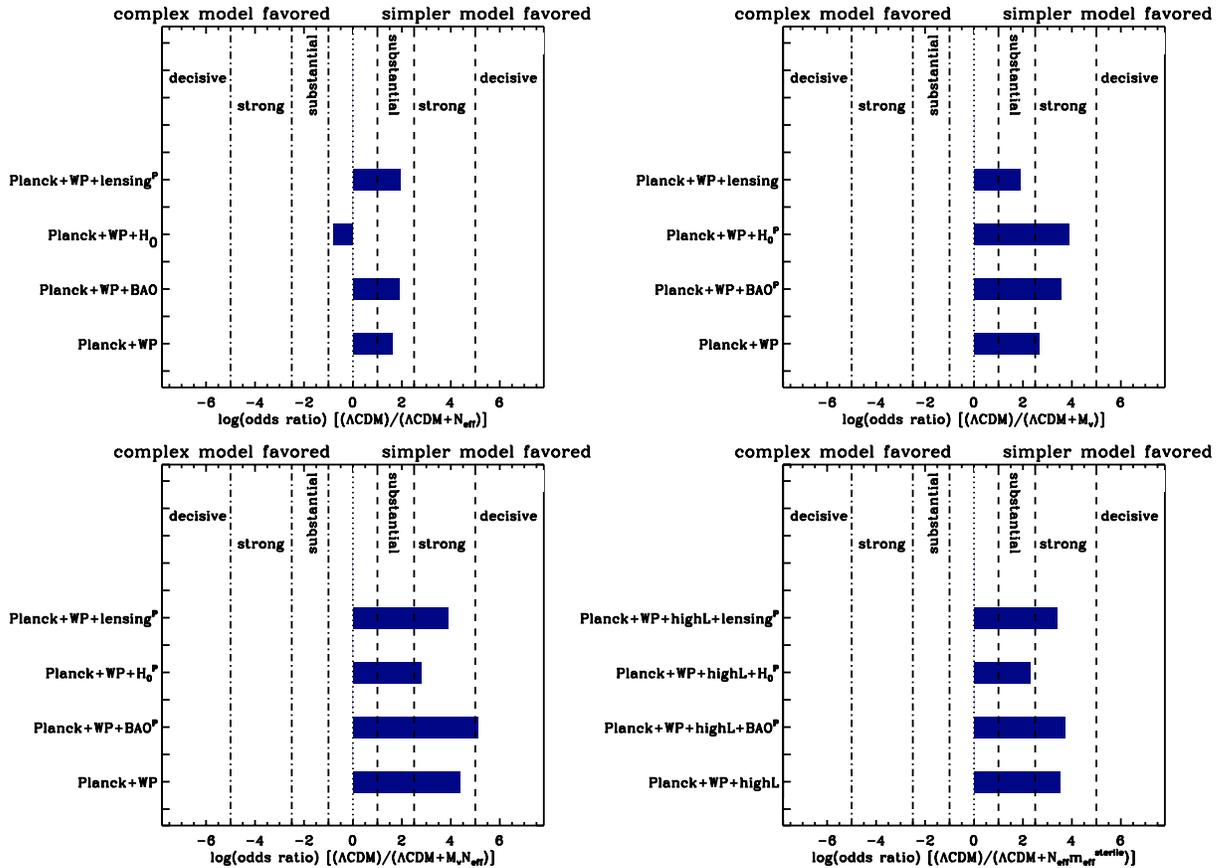


Figure 1. Summary of the Evidence ratios for the models discussed in the text for selected dataset combinations. The Evidence for the simplest model (Λ CDM) is substantial to strong, except when the “discrepant” H_0 measurement is added. In this case there is still no evidence to favor the more-complex Λ CDM+ N_{eff} model.

good fit to the data, and that more complex models worsen the fit, on average, over the prior ranges spanned by the extra parameters.

3.2 Profile likelihood ratios and comparison with Evidence

We compute the PLR for the Λ CDM+ N_{eff} model as outlined in Sec. 2.3, using the same dataset combinations as for the Evidence calculations (see Table 4). As with the Evidence ratios, the estimates of the PLRs from MCMCs are affected by sampling and binning errors. To estimate the sampling error on the PLR, we first find the difference between the maximum Λ CDM likelihood and the maximum Λ CDM+ N_{eff} likelihood in a small bin of width 0.01 centered on $N_{\text{eff}} = 3.046$.⁴ By weighting this difference by the square root of the ratio of the number of samples in each PLR bin to the number of samples in the small $N_{\text{eff}} \simeq 3.046$ bin, we obtain the errors on the binned PLR. We ensure that the PLR is robust to reasonable changes in the bin size.

The resulting PLR curves are plotted in Fig. 2 for a subset of the data combinations considered. The increase in discriminatory power represented by the *Planck* data is clear to see through comparison of this figure with its pre-*Planck* equivalent (Fig. 7 in Ref. [8]). We find that the PLR at the standard value of N_{eff} is within errors of its maximum for all dataset combinations apart from those involving the “discrepant” H_0 measurement, and even in this case the preference for non-standard N_{eff} is less than 2 “effective” standard deviations. This corroborates, in a prior-independent manner, the

⁴Where Λ CDM chains are available. Where they are not available, we substitute a typical value for this error, 0.7.

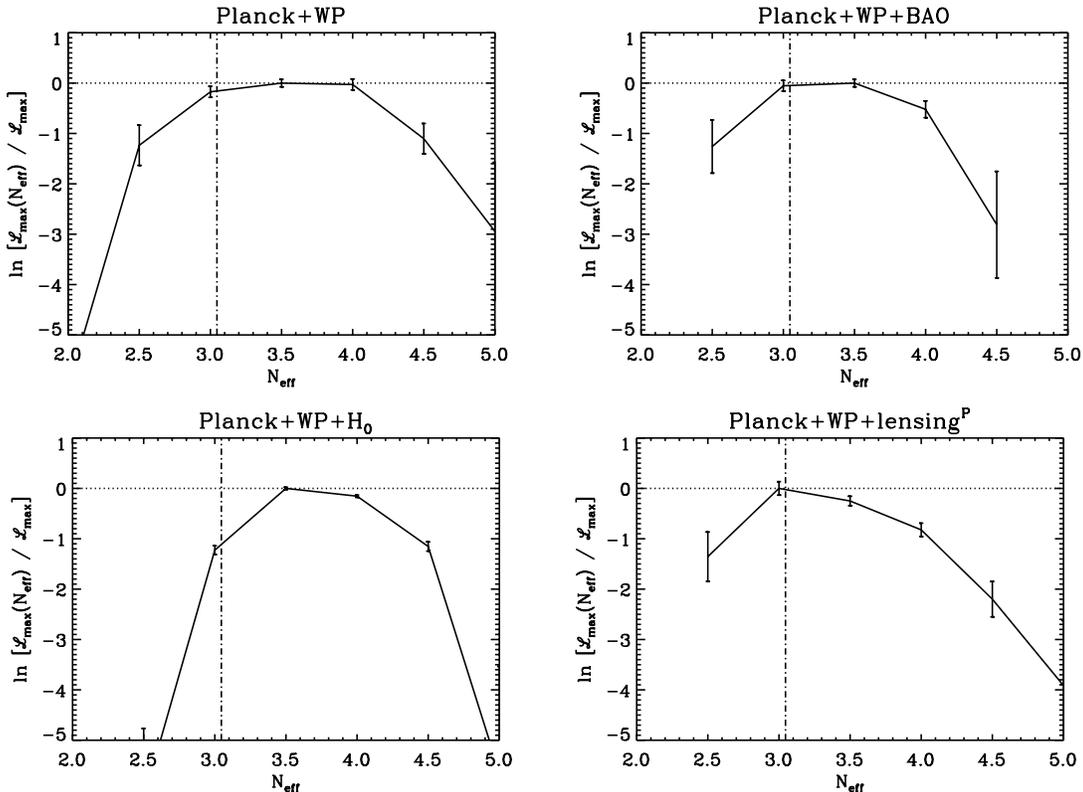


Figure 2. Profile likelihood ratio for the *Planck* “Gold” set alone and combined with BAO, H_0 and lensing data. The error bars for the Gold and Gold+lensing sets are computed using the corresponding Λ CDM chain as described in the main text; errors must be estimated for the other two plots as Λ CDM chains are not available. The dot-dashed vertical lines show the standard N_{eff} value of 3.046.

inference from the Evidence that the simpler Λ CDM model is preferred compared to a model with non-standard neutrino properties.

The PLR results are consistent with the Evidence findings: the maximum-likelihood parameters within the extended model are not significantly different from the Λ CDM values. The PLR also shows that, in agreement with the analysis of the posterior [13], values of $N_{\text{eff}} \sim 4$ cannot be excluded, but $N_{\text{eff}} \sim 5$ is disfavored. This is *not* in conflict with the Evidence findings: if we had a compelling reason to add one more (effective) neutrino to our base model (i.e., to fix N_{eff} to $3.046 + 1 = 4.046$), the Evidence would still mildly favor this simpler model over a model in which N_{eff} is a free parameter (albeit with $\ln(E_1/E_2) \lesssim 2$, and in many cases below 1). It is only if the base model had 5 neutrino species that the Evidence would favor the addition of N_{eff} as a parameter of the model. In this case, many data combinations (in fact, all combinations involving *Planck*+WP), would indicate *strong* and even *highly significant* Evidence for the Λ CDM+ N_{eff} model.⁵

3.3 Exploring degeneracies with other interesting parameters: the case of n_s

It is well known that the parameters describing neutrino properties are, at least in part, degenerate with other cosmological parameters. The effect of the degeneracy with Y_P , e.g., is clearly described

⁵For the Λ CDM+ Y_P + N_{eff} model the evidence is slightly weaker; however, for the *Planck*+WP+BAO, *Planck*+WP+lensing, and *Planck*+WP+highL+BAO data combinations, the Evidence for a departure of N_{eff} from the fiducial value would still be *strong* and even *highly significant*. Similarly, we would have to assume a fixed $M_\nu = 0.5$ eV in the fiducial model for the Evidence to indicate (at a *strong* or *highly significant* level) the need for M_ν as an extra parameter for at least some of the data combinations (*Planck*+WP+BAO/ H_0 , *Planck*+WP+lensing+BAO/ H_0 and *Planck*+WP+highL+BAO/ H_0).

data	evidence ratio: $\ln(E_{\Lambda\text{CDM}}/E_{\text{HZ}+N_{\text{eff}}})$					
	base	+BAO	+ H_0	+SNLS	+Union2	+lensing
Planck+WP	0.95	3.23	-1.86	-0.11	0.66	1.70 P

Table 5. Evidence results for the comparison of ΛCDM to a model with a scale-invariant power spectrum and non-standard effective number of neutrino species, $\text{HZ}+N_{\text{eff}}$. As previously, datasets marked “P” indicate that the results have been obtained by importance sampling (post-processing) the MCMCs; the extra datasets are always added individually to the base combination. Positive numbers mean that the simpler model is preferred. The errors on the evidence ratios are in most cases about ± 0.02 , with maximum values of ± 0.1 for thinned, post-processed chains.

in Ref. [54], but the optical depth to reionization, τ , and, more importantly, the spectral slope of the power spectrum, n_s , also exhibit degeneracies with N_{eff} . The n_s case is particularly interesting. With pre-*Planck* data (see e.g., Ref. [55] and references therein) a scale invariant power spectrum ($n_s = 1$) is consistent at the 95% confidence level when N_{eff} is allowed to vary as a parameter of the model. Within the ΛCDM model, *Planck* data rule out exact scale invariance at more than 5σ , indicating strong support for the inflationary paradigm. However, as discussed in Ref. [56], a model with $n_s = 1$ and $N_{\text{eff}} \sim 4$ fits the post-*Planck* CMB data almost as well as ΛCDM , although ΛCDM provides a much better fit when BAO data are also considered. As extensively discussed in e.g., Refs. [13, 54], the BAO measurements (by constraining the expansion history) are sensitive to the matter density, while the CMB geometric degeneracy implies higher cold dark matter density for higher N_{eff} .

These conclusions are based on local values of the likelihood: the maximum likelihood for a model with fixed $n_s = 1$ and varying N_{eff} (denoted $\text{HZ}+N_{\text{eff}}$) is compared to the maximum likelihood for ΛCDM (with varying n_s and fixed $N_{\text{eff}} = 3.046$). The scatter in such measurements is typically fairly large, and the comparison of maximum likelihoods does not provide a self-consistent model-selection criterion; thus it is interesting to determine whether the Evidence can discriminate between the two models. As the $\text{HZ}+N_{\text{eff}}$ and ΛCDM models are both nested within $\Lambda\text{CDM}+N_{\text{eff}}$ (at $n_s = 1$ and $N_{\text{eff}} = 3.046$, respectively), it is possible to form the desired Evidence ratio $E_{\Lambda\text{CDM}}/E_{\text{HZ}+N_{\text{eff}}}$ through two applications of the SDDR, determining the Evidence ratios with respect to $\Lambda\text{CDM}+N_{\text{eff}}$ as follows:

$$\begin{aligned}
\frac{E_{\Lambda\text{CDM}}}{E_{\text{HZ}+N_{\text{eff}}}} &= \frac{\Pr(\mathbf{d}|\Lambda\text{CDM})}{\Pr(\mathbf{d}|\text{HZ}+N_{\text{eff}})} \\
&= \frac{\Pr(N_{\text{eff}}|\mathbf{d}, \Lambda\text{CDM}+N_{\text{eff}})}{\Pr(N_{\text{eff}}|\Lambda\text{CDM}+N_{\text{eff}})} \Bigg|_{N_{\text{eff}}=3.046} \frac{\Pr(n_s|\Lambda\text{CDM}+N_{\text{eff}})}{\Pr(n_s|\mathbf{d}, \Lambda\text{CDM}+N_{\text{eff}})} \Bigg|_{n_s=1}. \quad (3.1)
\end{aligned}$$

The results of applying the SDDR to the *Planck* chains are presented in Table 5. The Evidence ratios support the findings of the *Planck* Collaboration’s maximum-likelihood analysis. For the Gold dataset, the Evidence ratio is fairly agnostic: there is at most a slight preference for ΛCDM . When BAO or lensing data are added, however, the Evidence is much more strongly in favor of ΛCDM . As discussed above, BAO and lensing constrain the matter density, breaking the parameter degeneracy for CMB data between N_{eff} and cold dark matter. Interestingly, when HST data are considered there is substantial Evidence in favor of the $\text{HZ}+N_{\text{eff}}$ model, as N_{eff} , n_s and H_0 are all positively correlated, and high values of N_{eff} and n_s can therefore resolve the tension between the CMB prediction and local measurements of H_0 .

4 Conclusions and future avenues

We have used Bayesian model comparison to determine whether extensions to Standard-Model neutrino physics in the form of additional effective numbers of neutrinos and/or massive neutrinos are merited by the latest cosmological data. The official *Planck* analysis and subsequent analyses using *Planck* data have concentrated on parameter estimation, and have therefore not addressed the model comparison issue considered in this work. We find that the state-of-the-art cosmological data do not

show (Bayesian) Evidence for deviations from the standard Λ CDM cosmological model (which has three massless neutrino families), findings that are corroborated by a prior-independent analysis based on the PLR. This is particularly interesting for two reasons.

Firstly, perhaps the strongest evidence the *Planck* data have provided for inflation is ruling out exact scale-invariance at $\sim 5\sigma$. As the official *Planck* analysis acknowledges, this statement is model dependent: the strongest constraint derives from a comparison of vanilla Λ CDM with a model with a Harrison Zel’dovich power spectrum, but the constraint is significantly relaxed when the scale-invariant model contains additional effective neutrino species [13]. We have bolstered the *Planck* team’s maximum-likelihood-ratio investigation by determining whether the Bayesian Evidence prefers a model with a Harrison-Zel’dovich power spectrum and freely-varying N_{eff} over standard Λ CDM. Using only primary CMB data, the Evidence indicates that there is at most a slight preference for Λ CDM, but when BAO or CMB lensing data are added, this preference increases significantly. While there is considerable variation in the results for other dataset combinations, likely stemming from tensions with the Planck data, it appears that the clear inflationary prediction of near, but not exact, scale-invariance is confirmed by cosmological data even when additional relativistic degrees of freedom are allowed.

Secondly, when considering one-parameter extensions of the Λ CDM model described by the addition of an effective number of neutrino species N_{eff} , the pre-*Planck* cosmological data have been often interpreted as supporting the case for dark radiation (i.e., deviations from the standard number of $N_{\text{eff}} = 3.046$), in turn interpreted as evidence for the existence of one or more sterile neutrinos. A Bayesian model-selection analysis did not find significant evidence in favor of extended neutrino physics [8]; nor, however, was it decisively in favor of the Standard Model. Post-*Planck*, the data leave less freedom to the sterile-neutrino interpretation, but Bayesian parameter estimates still leave room for dark radiation. In the present work we find that the Evidence for the simpler Λ CDM model has *increased* with the improvement in observational data represented by *Planck*, both alone and in combination with other datasets. The Evidence for the simpler model is “substantial” in the Jeffreys’ scale for many dataset combinations. Even with the discrepant local Hubble constant measurement, there is no Evidence for deviations from the standard N_{eff} . The Evidence findings are corroborated by the *prior-independent* PLR results. The *Planck* Collaboration used the comparison of maximum likelihoods – a similar approach to the PLR – to determine any preference for non-standard N_{eff} in selected dataset combinations. Our PLR and Evidence results are in qualitative agreement with the findings of the official *Planck* analysis.

The next *Planck* data release is expected to significantly improve the constraints on N_{eff} due to the additional information contained in the polarization signal. Changes in N_{eff} leave their specific signature in the polarization as thoroughly explained in Ref. [57] (see their Figs. 5 and 6 for the effect on the angular power spectrum). A Fisher-based forecast for the Evidence was presented in Ref. [58], finding that $\Delta N_{\text{eff}} = 1$ could be distinguished at a level between “substantial” and “strong”, even considering a two-parameter extension in which both N_{eff} and M_ν are free parameters. In this work, *Planck* Blue Book performance [59] was assumed, and the Evidence was computed under the approximation of Gaussian posteriors [60], but this should not change the conclusions qualitatively.

In the standard Λ CDM model neutrinos are massless. This makes the model strictly incorrect, as neutrino oscillations show that neutrinos have a non-zero mass. For this reason, the *Planck* team fixed the summed neutrino masses to a value close to the minimum allowed by oscillations in their “base” Λ CDM model. We have considered whether current cosmological data exhibit any Evidence for this (well motivated) “paradigm shift”. We find that the Evidence does not favor the extension of Λ CDM to include non-zero neutrino mass: in fact, the Evidence for the standard Λ CDM model is “substantial” to “strong”, depending on the dataset combination, and is stronger than it was using pre-*Planck* data.

This does not conflict with the evidence from neutrino oscillations for non-zero neutrino mass: we consider only cosmological data, which implies that there is no need for an extra parameter, M_ν , to describe cosmological data. This only means that M_ν is too small to be cosmologically relevant, i.e., smaller than the uncertainties of current cosmological measurements. Future data are expected to have enough statistical power to detect the effect of non-zero neutrino masses with high significance, even if

M_ν is close to the minimum mass given by oscillation experiments (see, e.g., Refs. [61–64]). However, in the context of current cosmological data, the Bayesian Evidence indicates that the standard Λ CDM model does not require any of the extensions we have considered.

Acknowledgments

LV is supported by the European Research Council under the European Community’s Seventh Framework Programme FP7-IDEAS-Phys.LSS 240117. SMF is supported by STFC and a grant from the Foundational Questions Institute (FQXi) Fund, a donor-advised fund of the Silicon Valley Community Foundation on the basis of proposal FQXiRFP3-1015 to the Foundational Questions Institute. HVP is supported by STFC, the Leverhulme Trust, and the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 306478-CosmicDawn. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), part of the High Energy Astrophysics Science Archive Center (HEASARC). HEASARC/LAMBDA is a service of the Astrophysics Science Division at the NASA Goddard Space Flight Center. This work is based on observations obtained with *Planck* (<http://www.esa.int/Planck>), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada. The development of *Planck* has been supported by: ESA; CNES and CNRS/INSU-IN2P3-INP (France); ASI, CNR, and INAF (Italy); NASA and DoE (USA); STFC and UKSA (UK); CSIC, MICINN and JA (Spain); Tekes, AoF and CSC (Finland); DLR and MPG (Germany); CSA (Canada); DTU Space (Denmark); SER/SSO (Switzerland); RCN (Norway); SFI (Ireland); FCT/MCTES (Portugal); and PRACE (EU). This research was supported in part by the National Science Foundation under Grant No. PHY11-25915.

References

- [1] R. T. Cox, *Probability, frequency and reasonable expectation*, *Am.J.Th.Phys* **14** (1946) 1.
- [2] A. Jaffe, *H_0 and Odds on Cosmology*, *Astrophys.J.* **471** (Nov., 1996) 24, [astro-ph/].
- [3] M. P. Hobson, S. L. Bridle, and O. Lahav, *Combining cosmological data sets: hyperparameters and Bayesian evidence*, *Mon.Not.R.Astron.Soc* **335** (Sept., 2002) 377–388, [astro-ph/].
- [4] A. R. Liddle, *How many cosmological parameters?*, *Mon.Not.R.Astron.Soc* **351** (July, 2004) L49–L53, [astro-ph/].
- [5] D. Parkinson, P. Mukherjee, and A. R. Liddle, *Bayesian model selection analysis of WMAP3*, *Phys. Rev. D* **73** (June, 2006) 123523, [astro-ph/].
- [6] R. Trotta and A. Melchiorri, *Indication for Primordial Anisotropies in the Neutrino Background from the Wilkinson Microwave Anisotropy Probe and the Sloan Digital Sky Survey*, *Phys.Rev.Lett.* **95** (June, 2005) 011305, [astro-ph/].
- [7] R. Trotta, *Applications of Bayesian model selection to cosmological parameters*, *Mon.Not.R.Astron.Soc* **378** (June, 2007) 72–82, [astro-ph/].
- [8] S. M. Feeney, H. V. Peiris, and L. Verde, *Is there evidence for additional neutrino species from cosmology?*, *JCAP* **1304** (2013) 036, [arXiv:1302.0014].
- [9] **Planck Collaboration** Collaboration, P. Ade et al., *Planck 2013 results. I. Overview of products and scientific results*, arXiv:1303.5062.
- [10] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, *Phys.Rept.* **429** (2006) 307–379, [astro-ph/0603494].
- [11] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, and P. D. Serpico, *Relic neutrino decoupling including flavour oscillations*, *Nucl.Phys.B* **729** (Nov., 2005) 221–234, [hep-ph/05].
- [12] K. Abazajian, M. Acero, S. Agarwalla, A. Aguilar-Arevalo, C. Albright, et al., *Light Sterile Neutrinos: A White Paper*, arXiv:1204.5379.
- [13] **Planck Collaboration** Collaboration, P. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, arXiv:1303.5076.

- [14] S. Weinberg, *Goldstone Bosons as Fractional Cosmic Neutrinos*, *ArXiv e-prints* (May, 2013) [[arXiv:1305.1971](#)].
- [15] N. Said, E. Di Valentino, and M. Gerbino, *Dark Radiation after Planck*, [arXiv:1304.6217](#).
- [16] E. Di Valentino, A. Melchiorri, and O. Mena, *Dark Radiation candidates after Planck*, [arXiv:1304.5981](#).
- [17] C. Kelso, S. Profumo, and F. S. Queiroz, *Non-thermal WIMPs as "Dark Radiation" in Light of ATACAMA, SPT, WMAP9 and Planck*, [arXiv:1304.5243](#).
- [18] J. Mastache and A. de la Macorra, *Extra relativistic degrees of freedom without extra particles using Planck data*, [arXiv:1303.7038](#).
- [19] P. Di Bari, S. F. King, and A. Merle, *Dark Radiation or Warm Dark Matter from long lived particle decays in the light of Planck*, [arXiv:1303.6267](#).
- [20] M. Archidiacono, E. Giusarma, S. Hannestad, and O. Mena, *Cosmic dark radiation and neutrinos*, [arXiv:1307.0637](#).
- [21] J. B. et al. (Particle Data Group) *Phys. Rev. D* **86** (2012) 010001.
- [22] Skilling, J., *Nested Sampling*, *AIP Conference Proceedings* **735** (2004) 395405.
- [23] Skilling, J., *Nested Sampling for General Bayesian Computation*, *Bayesian Analysis* **1** (2006), no. 4 833–860.
- [24] A. R. Liddle, P. Stefano Corasaniti, M. Kunz, P. Mukherjee, D. Parkinson, and R. Trotta, *Comment on 'Tainted evidence: cosmological model selection versus fitting', by Eric V. Linder and Ramon Miquel (astro-ph/0702542v2)*, *ArXiv Astrophysics e-prints* (Mar., 2007) [[astro-ph](#)].
- [25] H. Jeffrey, *The theory of probability*. Oxford University Press, 1961.
- [26] R. Kass and A. E. Raftery, *Bayes factors*, *JASA* **90** (1995), no. 430 773–795.
- [27] J. Dickey, *The weighted likelihood ratio, linear hypotheses on normal location parameters*, *Ann. Math. Statist.* **42** (1971) 204–223.
- [28] I. Verdinelli and L. Wasserman, *Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio*, *Journal of the American Statistical Association* **90** (1995) 430.
- [29] A. O'Hagan and F. J., *Kendalls advanced theory of Statistics: Bayesian inference*. Arnold, London, 2004.
- [30] J. Marin and C. P. Robert, *On resolving the SavageDickey paradox*, *Electron. J. Statist.* **4** (2010) 643–654.
- [31] R. van Haasteren, *Bayesian evidence: can we beat MultiNest using traditional MCMC methods?*, [arXiv:0911.2150](#).
- [32] M. D. Weinberg, *Computing the Bayes Factor from a Markov Chain Monte Carlo Simulation of the Posterior Distribution*, *Bayesian Anal.* **7** (2012), no. 3 737–770.
- [33] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Ann.Math.Statist.* **9** (1938), no. 1 pp. 60–62.
- [34] B. A. Reid, L. Verde, R. Jimenez, and O. Mena, *Robust Neutrino Constraints by Combining Low Redshift Observations with the CMB*, *JCAP* **1001** (2010) 003, [[arXiv:0910.0008](#)].
- [35] A. X. Gonzalez-Morales, R. Poltis, B. D. Sherwin, and L. Verde, *Are priors responsible for cosmology favoring additional neutrino species?*, [arXiv:1106.5052](#).
- [36] C. Bennett, D. Larson, J. Weiland, N. Jarosik, G. Hinshaw, et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, [arXiv:1212.5225](#).
- [37] S. Das, T. Louis, M. R.olta, G. E. Addison, E. S. Battistelli, et al., *The Atacama Cosmology Telescope: Temperature and Gravitational Lensing Power Spectrum Measurements from Three Seasons of Data*, [arXiv:1301.1037](#).
- [38] C. Reichardt, L. Shaw, O. Zahn, K. Aird, B. Benson, et al., *A measurement of secondary cosmic microwave background anisotropies with two years of South Pole Telescope observations*, *Astrophys.J.*

- 755 (2012) 70, [arXiv:1111.0932].
- [39] **Planck collaboration** Collaboration, P. Ade et al., *Planck 2013 results. XV. CMB power spectra and likelihood*, arXiv:1303.5075.
 - [40] **WMAP Collaboration** Collaboration, G. Hinshaw et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, arXiv:1212.5226.
 - [41] **Planck Collaboration** Collaboration, P. Ade et al., *Planck 2013 results. XVII. Gravitational lensing by large-scale structure*, arXiv:1303.5077.
 - [42] **SDSS Collaboration** Collaboration, W. J. Percival et al., *Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Data Release 7 Galaxy Sample*, *Mon.Not.Roy.Astron.Soc.* **401** (2010) 2148–2168, [arXiv:0907.1660].
 - [43] N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, et al., *A 2Methods and Application to the Sloan Digital Sky Survey*, arXiv:1202.0090.
 - [44] C. Blake, E. Kazin, F. Beutler, T. Davis, D. Parkinson, et al., *The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations*, *Mon.Not.Roy.Astron.Soc.* **418** (2011) 1707–1724, [arXiv:1108.2635].
 - [45] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample*, *Mon.Not.Roy.Astron.Soc.* **427** (2013), no. 4 3435–3467, [arXiv:1203.6594].
 - [46] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, et al., *The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant*, *Mon.Not.Roy.Astron.Soc.* **416** (2011) 3017–3032, [arXiv:1106.3366].
 - [47] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, et al., *A 3Telescope and Wide Field Camera 3*, *Astrophys.J.* **730** (2011) 119, [arXiv:1103.2976].
 - [48] A. Conley, J. Guy, M. Sullivan, N. Regnault, P. Astier, et al., *Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova Legacy Survey*, *Astrophys.J.Suppl.* **192** (2011) 1, [arXiv:1104.1443].
 - [49] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, et al., *The Hubble Space Telescope Cluster Supernova Survey: V. Improving the Dark Energy Constraints Above $z=1$ and Building an Early-Type-Hosted Supernova Sample*, *Astrophys.J.* **746** (2012) 85, [arXiv:1105.3470].
 - [50] F. Feroz and M. Hobson, *Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis*, *Mon.Not.R.Astron.Soc.* **384** (2008) 449, [arXiv:0704.3704].
 - [51] F. Feroz, M. Hobson, and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*, *Mon.Not.R.Astron.Soc.* **398** (2009) 1601–1614, [arXiv:0809.3437].
 - [52] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. Nolta, et al., *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-Derived Parameters*, *Astrophys.J.Suppl.Ser.* **192** (2011) 16, [arXiv:1001.4635].
 - [53] R. Keisler, C. Reichardt, K. Aird, B. Benson, L. Bleem, et al., *A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope*, *Astrophys.J.* **743** (2011) 28, [arXiv:1105.3182].
 - [54] Z. Hou, R. Keisler, L. Knox, M. Millea, and C. Reichardt, *How Additional Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail*, arXiv:1104.2333.
 - [55] H. Li, J.-Q. Xia, and X. Zhang, *Constraints on Scalar Spectral Index from Latest Observational Measurements*, arXiv:1303.3428.
 - [56] **Planck Collaboration** Collaboration, P. Ade et al., *Planck 2013 results. XXII. Constraints on inflation*, arXiv:1303.5082.
 - [57] S. Bashinsky and U. Seljak, *Signatures of relativistic neutrinos in CMB anisotropy and matter clustering*, *Phys. Rev. D* **69** (Apr., 2004) 083002, [astro-ph/].
 - [58] T. D. Kitching, A. F. Heavens, L. Verde, P. Serra, and A. Melchiorri, *Finding evidence for massive*

- neutrinos using 3D weak lensing*, *Phys. Rev. D* **77** (May, 2008) 103008, [[arXiv:0801.4565](#)].
- [59] The Planck Collaboration, *The Scientific Programme of Planck*, *ArXiv Astrophysics e-prints* (Apr., 2006) [[astro-ph/](#)].
- [60] A. F. Heavens, T. D. Kitching, and L. Verde, *On model selection forecasting, dark energy and modified gravity*, *Mon.Not.R.Astron.Soc.* **380** (Sept., 2007) 1029–1035, [[astro-ph/](#)].
- [61] C. Carbone, L. Verde, Y. Wang, and A. Cimatti, *Neutrino constraints from future nearly all-sky spectroscopic galaxy surveys*, *JCAP* **1103** (2011) 030, [[arXiv:1012.2868](#)].
- [62] B. Audren, J. Lesgourgues, S. Bird, M. G. Haehnelt, and M. Viel, *Neutrino masses and cosmological parameters from a Euclid-like survey: Markov Chain Monte Carlo forecasts including theoretical errors*, *JCAP* **1301** (2013) 026, [[arXiv:1210.2194](#)].
- [63] T. Basse, O. E. Bjaelde, J. Hamann, S. Hannestad, and Y. Y. Y. Wong, *Dark energy and neutrino constraints from a future EUCLID-like survey*, [arXiv:1304.2321](#).
- [64] J. Hamann, S. Hannestad, and Y. Y. Wong, *Measuring neutrino masses with a future galaxy survey*, *JCAP* **1211** (2012) 052, [[arXiv:1209.1043](#)].