Accounting for Correlations When Fitting Extra Cosmological Parameters

Y. Huang ©, G. E. Addison ©, and C. L. Bennett ©
Dept. of Physics & Astronomy, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218-2686, USA; yhuang98@jhu.edu

Received 2019 April 23; revised 2019 July 3; accepted 2019 July 24; published 2019 September 10

Abstract

Current cosmological tensions motivate investigating extensions to the standard Λ cold dark matter (ΛCDM) model. Additional model parameters are typically varied one or two at a time, in a series of separate tests. The purpose of this paper is to highlight that information is lost by not also examining the correlations between these additional parameters, which arise when their effects on model predictions are similar even if the parameters are not varied simultaneously. We show how these correlations can be quantified with simulations and Markov Chain Monte Carlo methods. As an example, we assume that ΛCDM is the true underlying model, and calculate the correlations expected between the phenomenological lensing amplitude parameter, AL, the running of the spectral index, nrun, and the primordial helium mass fraction, Yp, when these parameters are varied one at a time along with the ΛCDM parameters in fits to the Planck 2015 temperature power spectrum. These correlations are not small, ranging from 0.31 (AL − nrun) to −0.93 (nrun−Yp). We find that the values of these three parameters from the Planck data are consistent with ΛCDM expectations within 0.9σ when the correlations are accounted for. This does not explain the 1.8–2.7σ Planck preference for AL > 1, but provides an additional ΛCDM consistency test. For example, if AL > 1 was a symptom of an underlying systematic error or some real but unknown physical effect that also produced spurious correlations with nrun or Yp, our test might have revealed this. We recommend that future cosmological analyses examine correlations between additional model parameters in addition to investigating them separately, one at a time.

Key words: cosmic background radiation – cosmological parameters – cosmology: observations

1. Introduction

Over the last decade, much progress has been made on putting precise constraints on cosmological parameters within the Λ cold dark matter (ΛCDM) model, particularly from Cosmic Microwave Background (CMB) experiments (e.g., Bennett et al. 2013;Sievers et al. 2013;Story et al. 2013;Planck Collaboration XIII 2016). In the most recent release of Planck results (Planck Collaboration VI 2018), determinations of the standard ΛCDM parameters such as baryon density, Hubble constant, and matter density have reached the percent level or below.

Although currently there is no convincing evidence for deviations from the standard ΛCDM model from any single experiment, tensions exist between the values of some parameters inferred from different data sets. The most severe one is the 4.4σ disagreement between the Hubble constant measurements from the anchor-Cepheid-supernova distance ladder by the SH0ES collaboration (Riess et al. 2019) and from the Planck CMB data (Planck Collaboration VI 2018). The measurement of H0 via strong lensing time delays (Bonvin et al. 2017; Birrer et al. 2019) is consistent with the distance ladder and in 2.5σ tension with Planck. Addison et al. (2018) showed that the tension between early and late time universe measurements persists even without the inclusion of Planck data, using baryon acoustic oscillation (BAO) scale measurements.

Extensions or alternatives to the ΛCDM model have been explored in an attempt to resolve the Hubble tension. For example, the effects of varying the effective number of neutrino species (e.g., Riess et al. 2016) and the equation of state parameter of dark energy (e.g.,Joudaki et al. 2017) have been studied, though these extensions have not been able to effectively relieve the tension without including multiple turning points in the evolution of the dark energy equation of state (Zhao et al. 2017).

Recently, new ideas have been proposed as more promising solutions, for example, with the introduction of early dark energy (Poulin et al. 2019) or self-interaction massive neutrinos (Kreisch et al. 2019).

If there is new physics, consistency tests within the ΛCDM model will eventually fail with sufficiently sensitive new data. An example of a parameter used to test consistency is the lensing amplitude, AL. It was first introduced by Calabrese et al. (2008) as a phenomenological way to quantify the effect of weak gravitational lensing in the CMB power spectrum. By definition, AL = 1 is the physical value. However, the Planck temperature power spectrum (TT) data have shown a persistent preference for AL > 1 at 1.8–2.7σ depending on which data sets are included (Planck Collaboration XVI 2014; Planck Collaboration XIII 2016; Planck Collaboration VI 2018). For discussion of the AL tension, see also Addison et al. (2016), Motloch & Hu (2018, 2019). The cause of the deviation of AL from its physical value is unclear, however, varying AL is an example of the sort of test that might ultimately shed light on the origin of the distance ladder tension.

Typically results from fitting additional model parameters are presented one or two at a time, a series of separate tests (for recent examples, see Heavens et al. 2017; Joudaki et al. 2017; Planck Collaboration VI 2018). In this paper, we emphasize that constraints on some of these extension parameters may be correlated. Information is lost when ignoring these correlations. For example, when different parameters have similar effects on a theory prediction they cannot be perfectly distinguished with some given set of data. Also, for example, if specific types of correlations between a pair of parameters are expected in a theory, that can be tested for and provide valuable additional information.

As an illustration, the red point in Figure 1 shows the maximum-likelihood (ML) values for two parameters, A and B,
obtained from separate $\Lambda$CDM+$A$ and $\Lambda$CDM+$B$ fits. The green contours show the expected distribution of $A$ and $B$ fitted separately (estimated from e.g., simulations) if $\Lambda$CDM, with $A = 0$ and $B = 0$, is the true model. For this example we do not fit $A$ and $B$ at the same time in a $\Lambda$CDM+$A+B$ model.

Only looking at the ML values one at a time, as is usually done, would lead to an incorrect conclusion of consistency, as the ML values of $A$ and $B$ are each only 1 standard deviation away from their fiducial $\Lambda$CDM values. In contrast, in the 2D space the shift from the fiducial point is orthogonal to the expected degeneracy direction and the ML point is actually outside the 99.7% contour. This shows that there is a strong disagreement between the data and the assumed $\Lambda$CDM model that was not revealed by treating $\Lambda$CDM+$A$ and $\Lambda$CDM+$B$ as independent tests. We are not concerned here with the cause of the disagreement (e.g., systematic errors, an incorrect model, etc.), but only with the loss of information from sequential fits of one independent parameter at a time. The point is that accounting for the correlation between the extension parameters can give additional information.

Therefore, to more carefully assess whether the standard $\Lambda$CDM model can consistently describe data, the covariance of the set of extension parameters should be quantified.

In this paper, our goal is to answer the following questions:

1. How do we calculate the expected correlation between different extension parameters to the standard $\Lambda$CDM model when constrained by the same data?
2. How do we incorporate these correlations into a more stringent test of the $\Lambda$CDM model?

As an example, we use the Planck 2015 temperature power spectrum likelihood code (Planck Collaboration XI 2016). The outline of this paper is as follows. In Section 2 we present the theoretical basis of our work and two different methods to achieve our goals. We show results in Section 3, followed by a discussion in Section 4 of general recipes for similar analysis in the future and conclusions in Section 5.

2. Methodology

2.1. Estimating Correlation between Extension Parameters Using Simulations

In general it is not straightforward to infer the correlation between parameters $A$ and $B$ directly from Markov Chain Monte Carlo (MCMC) fits to the $\Lambda$CDM+$A$ and $\Lambda$CDM+$B$ models that have already been performed by experimental collaborations like Planck. To make progress, we assume that the data likelihood is Gaussian, with a covariance matrix $\mathbf{C}$ that does not depend on the cosmological parameters, and that the posterior distribution for the cosmological parameters is also Gaussian.

The assumption of Gaussianity of the likelihood is made widely across different cosmological measurements, including CMB power spectra (e.g., Planck Collaboration XI 2016; Louis et al. 2017; Henning et al. 2018), weak lensing shear (e.g., Krause et al. 2017; Wright et al. 2018; Hikage et al. 2019), galaxy clustering (e.g., Percival et al. 2014; Alam et al. 2017), supernovae distance moduli (e.g., Scolnic et al. 2018; Abbott et al. 2019), and others. The likelihood can almost always be made more Gaussian through compression of the data, often with negligible loss of cosmological information (e.g., combining CMB power spectra over a range of multipoles into bins). Neglecting the cosmological parameter dependence in the covariance (e.g., assuming a fixed fiducial model and set of parameters for computing the cosmic variance contribution to the errors) has also been demonstrated to be a suitable approximation for Planck, and other current experiments (e.g., Hamimeche & Lewis 2008; Krause et al. 2017). The fiducial model as well as the covariance are often estimated from some iterative process of fitting the actual data.

For the Planck data, cosmological parameter posterior distributions can be well approximated as Gaussian for $\Lambda$CDM, as well as for many one- and two-parameter extensions, although there are also cases (e.g., involving curvature or varying the dark energy equation of state), where Planck alone does not provide Gaussian posteriors. Other experiments only constrain a portion of the $\Lambda$CDM parameter space sufficiently well to produce Gaussian constraints in one or two parameters. Many parameters are also physically required to be positive, truncating the available parameter space and causing departure from Gaussianity when the data do not constrain the parameter to significantly differ from zero. This is the case for current cosmological constraints on the neutrino mass, for example (see Figure 30 of Planck Collaboration XIII 2016). We return to the handling of non-Gaussian cases in Section 4.3.

In certain cases, the Bayesian posterior parameter distribution and the distribution of the ML parameter values from many realizations of the data are approximately equal. Specifically, $\mathbb{P}(\theta|d)$, the Bayesian posterior distribution of parameters $\theta$ sampled by MCMC given the experimental data, can well approximate $\mathbb{P}(\theta|d_{\text{sim}}(d_{\text{fid}}))|\theta_{\text{fid}}$. The latter is the distribution of frequentist ML parameter estimation based on realizations of $d_{\text{sim}}$ from a fiducial model $\theta_{\text{fid}}$. The choice of $\theta_{\text{fid}}$ is usually physically motivated and is based on fits from actual data. For a
detailed discussion, see e.g., Chapter 4 and Appendix B of Gelman et al. (2013).

We show in Appendix A that this correspondence is mathematically exact, if we: (1) assume the Gaussianity of both the data likelihood and the posterior distribution of parameters; (2) impose uninformative parameter priors that are far less constraining than the likelihood; (3) in the MCMC computation, replace the experimental data \( d \) with \( \mu(\theta_{\text{fid}}) \), the theory prediction of the fiducial model. In other words, \( \mathbb{P}(\theta | d = \mu(\theta_{\text{fid}})) \) from MCMC and \( \mathbb{P}(\theta | d = \mu(\theta_{\text{fid}})) \) from simulations are the same mathematically. In this work, the fiducial model is from fitting the Planck TT data. We will show in Section 3.2 that the exact choice of \( \theta_{\text{fid}} \) is not very important.

This correspondence allows us to estimate the correlation between extension parameters \( A \) and \( B \) using simulated data (i.e., frequency sampling of the likelihood) in the following steps:

1. Generate many simulated data sets (in our case simulated Planck-like CMB power spectra) drawn from the likelihood in the form of a Gaussian distribution \( \mathcal{N}(\mu(\theta_{\text{fid}}), \mathbf{C}) \) for some choice of fiducial ΛCDM model parameters, \( \theta_{\text{fid}} \), and using the covariance matrix \( \mathbf{C} \) provided by the experiment collaboration.
2. Calculate the maximum-likelihood parameters, \( \theta_{\text{ML}} \), for the ΛCDM+A and ΛCDM+B models for each simulated data set, and
3. Estimate the covariance between \( A \) in the ΛCDM+A fit and \( B \) in the ΛCDM+B fit using the sample covariance from the simulations.

2.2. Estimating Correlation between Extension Parameters using MCMC Chains

Alternatively, in the special case of Gaussian parameter posteriors, we can run MCMC chains to estimate the correlation between extension parameters directly. In Appendix B, we show that all the information on the correlation between extension parameters \( A \) and \( B \) can be estimated from three sets of chains: ΛCDM+A, ΛCDM+B, and ΛCDM+A+B. Specifically, we found that the correlation between \( A \) and \( B \) when they are fitted separately is equal to minus one times the correlation between \( A \) and \( B \) when they are fitted together. We refer to this property as correlation equivalence.

A complication of estimating correlation between extension parameters by performing MCMC on the real data in this way is that we clearly do not have control over the "underlying" cosmological model in the way we do when generating simulations. It is also possible that the real data have imperfections or systematic biases that are not correctly accounted for in the likelihood. We therefore instead performed MCMC computations replacing the real data vector with a theory prediction computed using the same \( \theta_{\text{fid}} \) as for the simulations.

2.3. An Example Using Planck CMB Spectra

Here we provide a worked example using Planck data for both the simulation and MCMC approaches discussed above. At the time of writing, the Planck 2018 likelihood code is not yet available. Therefore, we perform a simple three-parameter test using the Planck \( \ell \geq 30 \) 2015 TT data from the \texttt{plik_lite} likelihood\(^1\) (as described in Planck Collaboration XI 2016). This simplified likelihood includes only CMB information, marginalizing over foreground template amplitudes and other nuisance parameters.

2.4. Fiducial Model and Extension Parameters

We assume the fiducial model to be the best-fit ΛCDM model of \( \ell \geq 30 \) Planck TT \texttt{plik_lite} data with the optical depth fixed at \( \tau = 0.07 \), the Planck calibration parameter \( calPlanck \) equal to 1 and other nuisance parameters marginalized. The \( \ell < 30 \) likelihood is pixel based rather than power spectrum based. For simplicity we only include the power spectrum based likelihood of the \( \ell \geq 30 \) data. Moreover, because the \( \ell \geq 30 \) TT spectrum only well constrains the parameter combination \( A_e^{-2\tau} \), \( \tau \) is fixed to break the strong degeneracy with \( A_e \). As for the calibration parameter \( calPlanck \), a Gaussian prior with mean equal to 1 and standard deviation 0.0025 was originally imposed on it. Its variation has minimal impact on other parameters. For simplicity, \( calPlanck \) is fixed at 1.

For the standard ΛCDM parameters, we use

\[
\{\Omega_b h^2, \Omega_c h^2, 100\theta_MC, \tau, \log A_s, n_s\}_{\text{fid}} = \{0.02216, 0.1211, 1.0407, 0.07, 3.0777, 0.9601\}.
\]

In addition, the three extension parameters of interest and their fiducial values for this example are \( \{A_L, n_{run}, Y_P\}_{\text{fid}} = \{1, 0, 0.2453\} \). As mentioned in Section 1, \( A_L \) is a phenomenological parameter that artificially scales the lensing power spectrum. It is worth investigating as Planck has a curious preference for \( A_L > 1 \). Running of the spectral index, \( n_{run} \equiv d n_s / d \ln k \), and the primordial Helium mass fraction \( Y_P \) are an interesting pair of extension parameters to test because of their high correlation (~0.9), that is, they produce similar changes in the power spectrum damping tail. Even a small deviation from their degeneracy direction in their expected 2D distribution would be noticeable. In ΛCDM, \( Y_P \) is calculated from \( \Omega_b h^2 \) and the CMB temperature through big bang nucleosynthesis (BBN) predictions (Planck Collaboration XIII 2016). In ΛCDM+\( Y_P \), \( Y_P \) is independent of BBN and decouples from \( \Omega_b h^2 \). We impose flat priors on all model parameters with the default CosmoMC (Lewis & Bridle 2002) bounds.

The fiducial parameters from fitting the Planck data have uncertainties due to cosmic variance and experimental noise. A slightly different set of fiducial parameters, which are still consistent with the measured power spectrum, may produce different results in the covariance of the extension parameters as well as the significance of their deviations from the fiducial ΛCDM values. We found this effect to be small for the Planck 2015 data. See Section 3.2 for details.

2.5. Simulations

We run simulations to sample the distribution of the extension parameters. We draw 1000 binned power spectrum samples with mean equal to the binned power spectrum of the fiducial model and covariance equal to the \texttt{plik_lite} CMB band power covariance. We use the same binning scheme as Planck 2015 to convert power spectra computed by \texttt{CAMB}\(^2\) to band powers. We replace the data spectrum in the \texttt{plik_lite} likelihood with our samples, thus forming the simulated likelihoods. Next,

\(^1\) Can be downloaded from \texttt{http://pla.esac.esa.int/pla/#cosmology.}

\(^2\) \texttt{https://camb.info/}
using the ML finding algorithm in CosmoMC, setting \( \tau = 0.07 \) and the Planck calibration parameter \( \text{calPlanck} = 1 \), we maximize the simulated likelihoods to obtain the best-fit parameters of a specific model for each realization. The models we explore are: the standard \( \Lambda \text{CDM}, \Lambda \text{CDM}+A_L, \Lambda \text{CDM}+n_{\text{run}} \) and \( \Lambda \text{CDM}+Y_P \).

To quantify the overall shift between the values of the extension parameters estimated from the actual data and their fiducial \( \Lambda \text{CDM} \) values, we calculate the \( \chi^2 \) and its probability to exceed (PTE) for a \( \chi^2 \) distribution with degrees of freedom equal to the number of extension parameters:

\[
\chi^2 = (\theta_{\text{dat}} - \theta_{\text{fid}}) \Sigma_{\text{ex}}^{-1} (\theta_{\text{dat}} - \theta_{\text{fid}}),
\]

where \( \Sigma_{\text{ex}} \) is the covariance matrix for the extension parameters only, \( \theta_{\text{dat}} \) is an array of the extension parameters estimated from the actual data in one-parameter extensions and their fiducial values are statistically insignificant.

To verify that our priors on the parameters can be considered as uninformative and thus do not impact the degree of freedom for the \( \chi^2 \), we compute the \( \chi^2 \) values and their PTE assuming three degrees of freedom using the best-fit parameters of a specific model for all simulations. Then we compare the distribution of PTE values with the expected one with three degrees of freedom and find that they are consistent.

### 2.6. MCMC

As mentioned in Section 2.2, we can make use of correlation equivalence to estimate the expected covariance \( \Sigma_{\text{ex}} \) between extension parameters fitted separately by running MCMC chains on the data likelihood, with the experimental power spectrum replaced by the fiducial one. Again we set \( \tau = 0.07 \) and the Planck calibration parameter \( \text{calPlanck} = 1 \).

The diagonal elements in \( \Sigma_{\text{ex}} \) are estimated from the variances of the specific extension parameters from running the MCMC chains with the modified Planck \( \text{plik lite} \) TT likelihood, on the model \( \Lambda \text{CDM}+A_L, \Lambda \text{CDM}+n_{\text{run}} \) and \( \Lambda \text{CDM}+Y_P \).

For the off-diagonal elements, first we obtain the correlation between two extension parameters (again, denoting them generically as \( A \) and \( B \)) from the MCMC runs that vary both of them in the \( \Lambda \text{CDM}+A+B \) model with the same data set. Then we calculate the correlation between any two parameters, given their variances as described in the last paragraph:

\[
\text{Cov}(A, B)_{\text{fitted separately}} = -\text{corr}(A, B)_{\Lambda \text{CDM}+A+B} \times (\text{Var}(A)_{\Lambda \text{CDM}+A+B} \times \text{Var}(B)_{\Lambda \text{CDM}+A+B})^{1/2}.
\]

This way, we are able to estimate all of the elements in \( \Sigma_{\text{ex}} \) without running simulations at all, and can calculate the \( \chi^2 \) defined in Equation (1) and its PTE to quantify the shifts of the extension parameters from their fiducial values.

### 3. Results

#### 3.1. Quantifying Significance of Deviations of Extension Parameters from their Fiducial Values

In Table 1, we show the \( \chi^2 \) of the difference, as defined in Equation (1), their corresponding PTE values and the level of consistency with the assumed model in terms of the number of \( \sigma \). Our results imply no significant discrepancy (0.9\( \sigma \)) between the values of the three additional parameters estimated from the actual data in one-parameter extensions and their fiducial \( \Lambda \text{CDM} \) values. This is consistent with expectations if the standard \( \Lambda \text{CDM} \) is the correct model.

Taking a closer look at elements in the estimated covariance, we show in Table 2 the estimated uncertainties of one-parameter extensions to the base \( \Lambda \text{CDM} \) model, along with the estimated parameter correlations from the two methods. These results are consistent up to at most 5\%, in the \( Y_P \) uncertainty. The differences between using the two methods are likely due to numerical noise and have only a minimal impact on our main conclusion.

We also visualize the shifts of the ML extension parameters from their fiducial values compared with their expected distributions in Figure 2. Notice in the 2D contour plots, how the ML points lie along the correlation/degeneracy direction of the estimated distributions, as expected if the base \( \Lambda \text{CDM} \) is the true model. Looking only at single extensions, the discrepancy between the ML value of \( A_L \) inferred by the \( \text{plik lite} \) TT data and its fiducial value is only significant at 1.8\( \sigma \) (from the MCMC method) or 1.9\( \sigma \) (from simulations), while the deviations of \( n_{\text{run}} \) and \( Y_P \) from their fiducial values are only significant at 0.8\( \sigma \) and 0.7\( \sigma \), respectively, from both running MCMC and simulations.

We do not attempt to explain or deemphasize the tension in \( A_L \). Rather, we note that there is no extra sign of discrepancy when we take into account its correlation with \( n_{\text{run}} \) or \( Y_P \). The significance of the \( \chi^2 \) of difference for the three parameters is not only reduced by the very small deviations of \( Y_P \) and \( n_{\text{run}} \) but is also reduced because the Planck \( \text{plik lite} \) TT data prefers the shifts in the parameters along their expected degeneracy directions. Considering the \( A_L- n_{\text{run}} \) or the \( A_L-Y_P \)
pair for example, their correlations are nonnegligible: 0.31 and −0.46, respectively. This information is not contained in the single extension tests. Table 1 and Figure 2 together show that the correlation of \( A_L \) with \( n_{\text{run}} \) and \( Y_P \) are as expected if \( \Lambda \) CDM is the true model, even though \( A_L \) deviates moderately with its fiducial value. Had we found inconsistency between the ML parameters from the data and the expected \( \Lambda \) CDM joint distribution, it could indicate the presence of systematic error or unknown physical effects producing spurious parameter correlations.

In addition, numerical differences between the simulated contours and those of MCMC are small in this figure, again providing confidence to both methods.

3.2. Testing Stability of Results Against Uncertainties in Fiducial Model

As mentioned in Section 2.4, the fiducial parameters are from fitting the actual plik_lite data and therefore have uncertainties. So the standard \( \Lambda \) CDM model we define is not just one point in the parameter space, but an ensemble of points whose shape is described by the parameter posterior distribution inferred from the data.

To assess the impact of different fiducial values on our results, we randomly draw 1000 sets of parameters from the MCMC chains fitted to the plik_lite data, with the base \( \Lambda \) CDM model, \( \tau = 0.07 \) and calPlanck = 1. For each set of parameters, we approximate the parameter covariance matrix by computing the \( C_\ell \) derivatives and Fisher matrices (see Equation (14) in Appendix B). Then, we use correlation equivalence to calculate the extension parameter covariance \( \Sigma_{ex} \). We find that varying parameters causes less than 3\% scatter in the matrix elements.

We also plug \( \Sigma_{ex} \) into Equation (1) and find that the resulting \( \chi^2 \) ranges from 4.2 to 5.1 and the PTE from 0.17 to 0.24, corresponding to consistency within 0.7–1\( \sigma \). These results show that the uncertainties in the fiducial model have very minimal impact on results from our consistency test. The stability of the results reflects the constraining power of the Planck data on the model parameters in \( \Lambda \) CDM.

4. Discussion

In Sections 2 and 3 we have shown an example of quantifying the level of consistency between extension parameters and their fiducial values given a specific data set. In this section, we outline and discuss steps to perform this approach more comprehensively for the Planck 2018 and other cosmology data sets.
4.1. MCMC Method

In the ideal case where all extension parameters of interest are Gaussian, we can estimate their expected distribution from MCMC chains. We denote the individual extension parameter as $\theta_{ex,i}$ with $i$ runs over 1 to $n_{ex}$, the total number of extension parameters. The suggested recipe is as follows:

1. Define a fiducial model (e.g., a fit from existing data) and calculate its theory prediction.
2. In the data likelihood of interest, substitute the experimental observables, e.g., power spectra in the CMB case, by the fiducial prediction.
3. Explore parameter space around the fiducial point by running MCMC chains on the modified data likelihood (s), fitting models $\Lambda$CDM+$\theta_{ex,i}$ and $\Lambda$CDM+$\theta_{ex,i}+\theta_{ex,j}$.
4. Calculate the variance of $\theta_{ex,i}$ from the one-parameter extension, and the correlation between $\theta_{ex,i}$ and $\theta_{ex,j}$ from the two-parameter extension.
5. Using results from previous step, construct a covariance matrix $\Sigma_{ex}$ for $\theta_{ex}$, with the signs of all the correlations flipped.
6. Calculate $\chi^2$ defined in Equation (1) and its PTE to quantify deviation of experimental values to fiducial ones. Additionally, one can plot confidence ellipses such as in Figure 2 to visualize the deviations.
7. Check validity of the input fiducial model, for example by using a Fisher Forecast (e.g., Heavens 2009) to approximate parameter covariance and estimating the shifts in the $\chi^2$ when varying fiducial parameter values.

As an example, on a computer cluster with 12-core 2.5 GHz processors, three MCMC run with 8 parallel chains usually take $\lesssim$24 hr to converge for the 2015 prik_lite likelihood. The CPU time for running one MCMC job is at the order of 100 hr and the total computing time is $\sim$5 $\times$ 10^4 hr. For running all one-parameter and two-parameter extension fits for e.g., $n_{ex} = 11$, which is the number of extra parameters fitted in the publicly available 2015 Planck chains.

4.2. Simulations

We can use simulations as an alternative method to estimate the expected distribution of additional parameters around the fiducial point. As described in Sections 2.1 and 2.5, one can perform the following procedure:

1. Generate simulated observables of the fiducial model using data covariance.
2. For each simulation, estimate the best-fit $\Lambda$CDM+$\theta_{ex}^{t}$ model for each extension parameter.
3. Calculate the covariance of extension parameter $\Sigma_{ex}$, the $\chi^2$ of difference defined in Equation (1) and its PTE to quantify the significance of difference. Confidence ellipses can also be plotted.

For a rough estimate of computing time for the simulation method, using CosmoMC and the same computer cluster with 12 cores per CPU, we find that the running time of the best-fit-finding algorithm is approximately 1 hr. For $n_{ex} = 11$, $n_{sam} = 2000$ and running 4 parallel best-fit-finding jobs for each simulation (to reduce numerical noise and avoid obtaining results from local minima), the total computing time is $9 \times 10^3$ hr.

4.3. In the Case of Non-Gaussianity

To discuss the case of non-Gaussianity, we first need to clarify from what exactly non-Gaussianity arises. Recall that throughout this paper, we assume a Gaussian data likelihood and priors that are uninformative. If the data constrains the parameters well enough, changes in the model predictions can be treated as only linearly dependent on the parameters. This means that the Taylor expansion of the log likelihood around the maximum point is significant up to quadratic terms of the parameters and therefore the parameters are Gaussian (see Appendix A).

In short, non-Gaussianity of parameters is a result of the data not being constraining enough for fitting parameters. When this is the case, there may not be a mathematical correspondence between the ML parameter distribution from the simulations and the posterior distribution of the MCMC chains, as our argument in Appendix A depends on the assumption of Gaussianity. Besides, because our proof of correlation equivalence in Appendix B also rests upon approximations of the log likelihood to only the second order derivatives, the property now becomes questionable. This means that we cannot simply estimate the correlation between parameters fit separately from the MCMC chains where they are fitted together. What is worse is that the means and the covariance matrix no longer carry all information of the parameter distributions and one might need to evaluate high order tensors (Sellentin et al. 2014). Additionally, the $\chi^2$ test is inappropriate for non-Gaussian parameters and we need new ways (such as one proposed in Appendix C of Motloch & Hu 2019) to quantify the significance of the overall difference between the experimental parameters and their fiducial values.

An example of non-Gaussian parameters are those with priors that are more informative than the data, e.g., the neutrino mass, $\sum m_\nu$, and the tensor to scalar ratio, $r$, physically defined as non-negative. Current CMB data are not sufficiently constraining to pull these parameters away from lower bounds (BICEP2 Collaboration et al. 2018; Planck Collaboration VI 2018). Another non-Gaussian parameter given just the TT data, is the curvature density $\Omega_k$. Curvature can only be weakly constrained as allowing it to be free worsens the existing degeneracy between the physical matter density $\Omega_m$ and the Hubble constant $H_0$ (Zaldarriaga et al. 1997; Percival et al. 2002; Kable et al. 2019). In the MCMC method, a two-parameter extension model with both $\Omega_m$ and $\Omega_k$ results in very wide and non-Gaussian distributions, as they are highly correlated: a positive curvature has similar effect as an increase in the matter density. This means that we cannot simply estimate the correlation between parameters fit separately from the MCMC chains where they are fitted together. What is worse is that the means and the covariance matrix no longer carry all information of the parameter distributions and one might need to evaluate high order tensors (Sellentin et al. 2014). Additionally, the $\chi^2$ test is inappropriate for non-Gaussian parameters and we need new ways (such as one proposed in Appendix C of Motloch & Hu 2019) to quantify the significance of the overall difference between the experimental parameters and their fiducial values.

To reduce non-Gaussianity, one can always include more data sets if available to have greater constraining power on the parameter, e.g., include the BAO data (Alam et al. 2017) in parameter fitting along with Planck. However, when extensions are used as a means to test the internal consistency of one data set, adding extra data is not an option. Fortunately, there are existing methods of transforming non-Gaussian parameters into Gaussian ones. One such method is the Box–Cox transformation (Box & Cox 1964), Joachimi & Taylor (2011) and Schuhmann et al. (2016) applied this bijective transformation to non-Gaussian cosmological parameters. Theoretically, correlation equivalence still holds for the transformed Gaussian parameters. Thus we may still apply the procedure outlined in

---

3 Our computation was conducted on the computer cluster of the Maryland Advanced Research Computing Center. See https://www.marcc.jhu.edu/cyberinfrastructure/hardware/ for descriptions of its system architecture.
Section 4.1 for transformed parameters, obtain the expected distribution for transformed ones, and then transform them back to the model parameters. One thing to note is that the Box–Cox transformation does not guarantee Gaussianity. So one should check for the Gaussianity of transformed parameters, e.g., calculate the skewness and kurtosis of the resulting distribution, and if needed, apply a second transformation. For consistency, it is also a good idea to compare the resulting distribution of model parameters to simulations, or compare the covariance of the transformed parameters with predictions from Fisher Forecast, keeping in mind that Fisher Forecast assumes Gaussianity.

Further understanding of non-Gaussian scenarios is left for future efforts.

5. Conclusions

In this paper, we have presented a method to help identify potential new physics and/or systematic errors by calculating the correlations between additional parameters and performing a \( \Lambda CDM \) consistency test accounting for them. Usually extension parameters are added to the model separately, one at a time. However, different parameters may affect the theory prediction in a similar way, which means their values from a given data set can be correlated, even when they are fit separately. Examining the consistency of extension parameters with \( \Lambda CDM \) expectations, accounting for these correlations, provides an additional test of the model beyond looking at the results from individual one-parameter extensions.

Under the assumption of Gaussianity of both the likelihood and the posterior distribution of the parameters, one can fit a series of one-parameter extension models to simulated data and obtain the multivariate distribution of the extension parameters. With the base parameters marginalized over, the \( \chi^2 \) of difference and its PTE can be calculated to quantify the significance of deviations from \( \Lambda CDM \).

A more computationally economic alternative is to run MCMC, fitting the same series of one-parameter extension models and additionally two-parameter extensions across all the possible pairs in the set of parameters of interest. Using correlation equivalence as proven in Appendix B, the covariance matrix of the extension parameters fitted separately can be estimated from the results of these MCMC runs and so the expected multivariate Gaussian distribution can be obtained.

In an attempt to narrow the causes of the \( A_L \) anomaly in the Planck data and to possibly clarify the tensions between cosmological measurements, we looked at the example parameter combination \( \{ A_L, n_{\text{run}}, Y_P \} \), fitted with the Planck 2015 \( \text{plik}\_\text{lite} \) \( \ell \geq 30 \) TT data, as we do not yet have access to the Planck 2018 likelihood. Results from MCMC and simulations show that the deviations of the three additional parameters from their fiducial \( \Lambda CDM \) values are consistent with statistical fluctuations within 0.9\( \sigma \) when correlations are accounted for.

Although the cause of the reported 1.8–2.7\( \sigma \) preference (depending on the the specific combination of data sets) for \( A_L > 1 \) by the Planck CMB data is yet to be understood, we find no further evidence for discrepancy when considering the correlations between \( A_L \), \( n_{\text{run}} \), and \( Y_P \). This is not a trivial test, as the correlations are significant: approximately 0.31 for \( A_L - n_{\text{run}} \), –0.46 for \( A_L - Y_P \), and –0.93 for \( n_{\text{run}} - Y_P \). If the unphysical \( A_L > 1 \) is a symptom of an underlying systematic error or some real but unknown physical effect that also produced spurious correlations with \( n_{\text{run}} \) or \( Y_P \), our test could have revealed this.

We also tested the stability of results against the uncertainties in the parameters of the experimentally fitted fiducial model, and find that the change of the fiducial model has no impact on our conclusions, only shifting the PTE values from 0.17 to 0.24.

Furthermore, we discussed how our procedures depend on the assumption of Gaussianity of parameters. If the assumption is not valid, MCMC runs cannot simply be used to estimate correlations between parameter fitted separately, nor may there be a mathematical correspondence between parameter distributions from the simulations and the MCMC runs. Therefore, efforts might attempt to include Gaussianization of non-Gaussian parameters, such as using the Box–Cox transformation (Box & Cox 1964).

The procedures developed in this paper can and should be applied to more extensive lists of extension parameters, with existing and future cosmological data, to provide a more stringent test and complete view of \( \Lambda CDM \) consistency.

This research was supported in part by NASA grants NNX16AF28G and NNX17AF34G. This work was partly 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. This research project was conducted using computational resources at the Maryland Advanced Research Computing Center (MARCC). The GetDist Python package was used to make Figures 1 and 2. We would like to thank Josh Kable for providing his code for Fisher matrix calculations, Haoyu Wang for his help with the derivation in Appendix B, and Janet Weiland for her comments on a draft of this paper.

Appendix A

Mathematical Correspondence between Frequentist Maximum-likelihood Parameters and Bayesian Parameter Posterior

In this section, we will show that there is a mathematical correspondence between frequentist ML parameters and Bayesian parameter posterior, under conditions described below. In other words, \( P(\theta|d,\theta^{\text{fid}}) \), the distribution of ML parameters estimated from realizations of a fiducial model is mathematically the same as \( P(\theta|d = \mu(\theta^{\text{fid}})) \), the Bayesian posterior distribution of parameters given a set of data that matches the theory prediction of the same fiducial model.

Given a set of fiducial parameters, \( \theta^{\text{fid}} \), we can calculate its theory prediction \( \mu(\theta^{\text{fid}}) \). With \( C \) as the data covariance estimated experimentally, we can then draw realizations of the data vector \( d \) from a multivariate Gaussian distribution \( N(\mu(\theta^{\text{fid}}), C) \)

Up to some constants, the log probability density function of \( d \) is given by

\[
\log P(d^{\text{sim}}|\theta^{\text{fid}}) = -\frac{1}{2}(d^{\text{sim}} - \mu(\theta^{\text{fid}}))^T C^{-1}(d^{\text{sim}} - \mu(\theta^{\text{fid}})).
\]

(3)

For a given set of simulated data, \( d^{\text{sim}} \), we can estimate the set of parameters that best describe it by maximizing the log
likelihood
\[
\log L(\mathbf{d}) \equiv \log \mathbb{P}(d^{\text{sim}}|\mathbf{d}). \tag{4}
\]

The ML parameter estimates, \(\theta^{\text{ML}}\), are defined as the solutions to the simultaneous equations
\[
\frac{\partial \log \mathbb{P}(d^{\text{sim}}|\mathbf{d})}{\partial \theta_i} \bigg|_{\theta=\theta^{\text{ML}}} = 0, \tag{5}
\]
where the index \(i\) runs over all parameters. Taylor expanding around \(\theta^{\text{ML}}\) gives
\[
\log \mathbb{P}(d^{\text{sim}}|\mathbf{d}) = \log \mathbb{P}(d^{\text{sim}}|\mathbf{d})_{|\theta=\theta^{\text{ML}}} + \frac{1}{2} \sum_i \sum_j \left(\theta_i - \theta_i^{\text{ML}}\right) \left(\frac{\partial^2 \log \mathbb{P}(d^{\text{sim}}|\mathbf{d})}{\partial \theta_i \partial \theta_j}\right)_{|\theta=\theta^{\text{ML}}} \left(\theta_j - \theta_j^{\text{ML}}\right) + \mathcal{O}(\theta - \theta^{\text{ML}})^3.
\]

If higher-order terms can be neglected, Equation (6) shows that given a set of \(d\), parameters around the ML point can be approximated by a Gaussian distribution, with the covariance \(\Sigma\) given by
\[
\Sigma(\theta|d^{\text{sim}})_{ij} = \left[-\frac{\partial^2 \log \mathbb{P}(d^{\text{sim}}|\mathbf{d})}{\partial \theta_i \partial \theta_j}\right]_{|\theta=\theta^{\text{ML}}}^{-1}.
\]

from all simulated data \(d\). In order to emphasize its dependence on \(d\), we will write \(\Sigma(d|d^{\text{sim}})\) instead of \(\Sigma(d|\theta^{\text{ML}})\). The parameter means are the expectation of the posterior parameter distribution is given by Bayes’ theorem:

\[
\mathbb{P}(\theta|d) = \frac{\mathbb{P}(d|\theta)\mathbb{P}(\theta)}{\mathbb{P}(d)} \propto \mathbb{P}(d|\theta)\mathbb{P}(\theta), \tag{9}
\]

where \(\mathbb{P}(\theta)\) is the prior probability and \(\mathbb{P}(d)\) is the evidence. If the posterior is Gaussian we can write, again up to some constants,
\[
\log \mathbb{P}(\theta|d) = \frac{1}{2}(\theta - \bar{\theta})^T \Sigma^{-1}(\theta - \bar{\theta}), \tag{10}
\]

where \(\Sigma\) is the parameter covariance matrix given by (7). For flat priors \(\bar{\theta} = \theta^{\text{ML}}(d)\). If \(\theta^{\text{fid}}\) mentioned above is close to \(\theta^{\text{ML}}(d)\), then (8) and (10) is only different by a small offset. To make them exactly equal, we can choose to replace \(d\) with a theory prediction computed using \(\theta^{\text{fid}}\) so that \(\theta^{\text{ML}}\) becomes equal to \(\theta^{\text{fid}}\) and the Bayesian posterior now describes the distribution of parameters around the fiducial value:
\[
\mathbb{P}(\theta|d = \mu(\theta^{\text{fid}})) = \log \mathbb{P}(\theta|\theta^{\text{fid}}) = -\frac{1}{2}(\theta - \theta^{\text{fid}})^T \Sigma^{-1}(\theta - \theta^{\text{fid}}), \tag{11}
\]
which matches exactly the distribution of the frequentist ML parameter estimates given in (8). This is asymptotically true even for non-flat priors when the data are sufficiently constraining, and we do not discuss the exact choice of priors further here. For more information, we again direct the reader to, for example, Chapter 4 of Gelman et al. (2013). We note that if the priors are informative this equivalence is broken and the method described in Section 2.2 would not be valid.

### Appendix B

**Correlation Equivalence**

In this section, using the mathematical correspondence from Appendix A, we show that when using ML estimation, the correlation between two parameters varied separately (e.g., the correlation between parameters \(A\) and \(B\) in model \(\Lambda\text{CDM}+A\) and \(\Lambda\text{CDM}+B\)) is the same but with an opposite sign as the correlation between the same two parameters varying together (e.g., \(\Lambda\text{CDM}+A+B\)).

#### B.1. ML Estimation and Parameter Covariance

With \(L \equiv \log L(\mathbf{d})\), the log likelihood of parameters given a specific sample of \(d\) can be written as
\[
L = -\frac{1}{2}(d - \mu(\theta))^{T} C^{-1}(d - \mu(\theta)), \tag{12}
\]
up to some constants. \(\mu(\theta)\) is the theory prediction of the data given some set of parameters \(\theta\). For any \(d\), the goal is to find the \(\theta\) that maximize \(L\).

As in Appendix A, at ML, the partial derivative of the log likelihood with respect to (w.r.t.) \(\theta_i\) is zero (\(i\) runs over \(n\), the number of parameters). That is
\[
\frac{\partial L}{\partial \theta_i} \bigg|_{\text{ML}} = \frac{\partial L}{\partial \mu} \bigg|_{\text{ML}} \frac{\partial \mu}{\partial \theta_i} \bigg|_{\text{ML}} = 0. \tag{13}
\]

Close to the ML point in parameter space, we can Taylor expand the log likelihood to the second order in \(\Delta \theta_i \equiv \theta_i - \theta_i^{\text{ML}}\), as shown in Equation (6), where \(\theta_i\) is well described by a multivariate Gaussian distribution. The parameter means are the values that maximize the likelihood.

The parameter covariance \(\Sigma\) can be calculated as the inverse of the Fisher matrix \(F^{-1}\), with the Fisher matrix, estimated from MCMC chains or calculated analytically, defined as
\[
F_{ij} = \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \bigg|_{\text{ML}} = \mu_{ij}^{\text{ML}} C^{-1} \mu_{ij}^{\text{ML}}. \tag{14}
\]

From now on we use the comma notation to denote partial derivatives w.r.t. \(\theta\), that is, \(\mu_{ij} \equiv \frac{\partial \mu}{\partial \theta_i \partial \theta_j}\).

In the following proof, we continue to work under the assumption of Gaussianity for \(\theta\) and expand \(\mu\) only to the first
order of $\theta$, that is, assuming the Gaussian linear model (Raveri & Hu 2019), around a chosen fiducial model $\theta_{\text{fid}}$:

$$
\mu(\theta) \approx \mu_{\text{fid}} + \sum_{i=1}^{n} \mu_{i,\text{fid}}(\theta_i - \theta_{i,\text{fid}}),
$$

where we define $\mu_{i,\text{fid}} = \mu(\theta_{i,\text{fid}})$. In the Gaussian linear model, the Jacobian $\mu_i$ may be taken as constant over changes in $\theta$ and so $\mu_{i,\text{fid}}$ is approximately equal to $\mu_{i,\text{ML}}$. For simplicity, from here on, we just drop all the superscripts for $\mu_i$. In practice, the fiducial model is usually based on the ML model given by the data and updated iteratively if necessary.

Next, we move on to calculate the elements in parameter covariance $\Sigma$:

$$
\Sigma_{ij} = (F^{-1})_{ij},
$$

(16)

Remember that for any invertible square matrix $M$, its inverse can be calculated as

$$
(M^{-1})_{ij} = (-1)^{i+j} |M|_{ij} / |M|,
$$

(17)

where $|M|$ is the determinant of $M$. We use the notation "$M_{ij}$" to denote a smaller matrix, corresponding to $M$ with the $i$th row and the $j$th column removed (the backslash symbol is borrowed from the notation for set difference in set theory). Then $|M_{ij}|$ is a minor of $M$. The determinant for an $n \times n$ matrix $M$ can be written as

$$
|M| = \sum_{i=1}^{n} (-1)^{i+j} M_{ij} |M_{i\backslash j}|,
$$

(18)

for any $1 \leq j \leq n$.

Therefore,

$$
\Sigma_{ij} = (-1)^{i+j} |F_{i\backslash j}| / |F|.
$$

(19)

If we choose to fix the $i$th parameter, we can simply remove the $i$th row and column from the Fisher matrix, and calculate a new parameter covariance matrix from the revised Fisher matrix.

### B.2. Correlation between Parameters A and B, Varying Together

When two parameters, $A$ and $B$, are both variables, we can calculate their covariance from the Fisher matrix that includes them along with the base $\Lambda$CDM parameters. So their correlation is just

$$
\text{corr}(A, B) = \frac{\Sigma_{AB}}{(\Sigma_{AA} \Sigma_{BB})^{1/2}}.
$$

(20)

From here on, for clarity we use $A$ and $B$ as subscripts for the rows and columns in the matrix for the two parameters of interest.

Following from Equations (20), (21) becomes

$$
\text{corr}(A, B) = \frac{-|F_{A\backslash B}|}{(|F_{A\backslash A}| |F_{B\backslash B}|)^{1/2}}.
$$

(21)

The minus sign is from setting the index of $B$ equal to that of $A$ plus one.

### B.3. Correlation between Parameters A and B, Varying Separately

For a generic set of parameters, given the data array $d$ (experimental or simulated) and the Fisher matrix $F$, we can substitute (16) into (13) to obtain a relationship between the ML parameters $\theta_{\text{ML}}$ and the fiducial ones:

$$
\theta_{\text{ML}}(d) - \theta_{\text{fid}} = F^{-1} \mu^T C^{-1} (d - \mu).
$$

(22)

So we can express $\theta_{\text{ML}}$ as:

$$
\theta_{\text{ML}}(d) = F^{-1} \mu^T C^{-1} d + \hat{\theta},
$$

(23)

where we use the vector $\hat{\theta}$ to represent all the terms in (22) that are independent of data $d$:

$$
\hat{\theta} \equiv -F^{-1} \mu^T C^{-1} \mu_{\text{fid}} + \theta_{\text{fid}}.
$$

As we will show below, $\hat{\theta}$ does not contribute to the correlation between $A$ and $B$.

Again for clarity, we use $A$ and $B$ to index the rows and columns for parameters $A$ and $B$, respectively, while using the generic $i$ and $j$ to index $n$ base $\Lambda$CDM parameters. So $A$ is the $(n + 1)$th parameter and $B$ is the $(n + 2)$th.

When fixing $B$ to its fiducial $\Lambda$CDM value $B_{\text{fid}}$, the expression for the best estimate for other parameters is almost the same as Equation (23), except that $F \rightarrow F_{B\backslash B}$ and the elements associated with $B$ in $\mu_{i,\text{ML}}$, $\theta_{\text{ML}}$ and $\hat{\theta}$ are deleted. The expression for the rest of elements in $\theta_{\text{ML}}$:

$$
\begin{pmatrix}
\theta_{1,\text{ML}} \\
\theta_{2,\text{ML}} \\
\vdots \\
\theta_{n,\text{ML}}
\end{pmatrix} = (F_{B\backslash B})^{-1}
\begin{pmatrix}
\mu_{1}^T C^{-1} d \\
\mu_{2}^T C^{-1} d \\
\vdots \\
\mu_{n}^T C^{-1} d
\end{pmatrix} + \hat{\theta}.
$$

(24)

The last row of (25) gives

$$
A_{\text{ML}}(d) = \sum_{i=1}^{n} (F_{B\backslash B})^{-1}_{ii} \mu_{i}^T C^{-1} (d + \hat{\theta}).
$$

(25)

Similarly, fixing $A$ and letting $B$ vary leads to

$$
B_{\text{ML}}(d) = \sum_{i=1}^{n} (F_{A\backslash A})^{-1}_{ii} \mu_{i}^T C^{-1} (d + \hat{\theta}).
$$

(26)

The variance of $A$ and $B$ can be obtained from the Fisher matrix (14) and its minors (18), as follows:

$$
\text{Var}(A_{\text{ML}}) = (F_{B\backslash B})^{-1}_{AA} = \frac{|F_{B\backslash B}|}{|F_{A\backslash A}|}
$$

(27)

$$
\text{Var}(B_{\text{ML}}) = (F_{A\backslash A})^{-1}_{BB} = \frac{|F_{A\backslash A}|}{|F_{A\backslash A}|}.
$$

(28)
And the covariance between $A^{ML}$ and $B^{ML}$ is:

$$\text{Cov}(A^{ML}, B^{ML}) = (\langle A^{ML} \rangle - \langle A^{ML} \rangle)(\langle B^{ML} \rangle - \langle B^{ML} \rangle),$$  \hspace{1cm} (29)

with the brackets here representing the averaging over all realizations of the data $d$.

Recall that the data covariance $C$ is assumed to be fixed and with our assumption of a Gaussian Linear Model, all partial derivatives w.r.t. the parameters are also constant. Then in (26) and (27), only $d$ is a variable. Inserting (26) and (27) into (30), we find that terms involving $\delta_i$ and $\delta_j$ cancel. In addition, all constant terms can be taken out of the brackets, leaving only $(\langle d - \langle d \rangle \rangle)^2$, which is just $C$. This simplifies to

$$\text{Cov}(A^{ML}, B^{ML}) = \sum_{i,j=1}^{n} (\mathcal{F}_{B|B})^{-1}_{ij} \mu_i^T C^{-1} \mu_j (\mathcal{F}_{A\setminus\Lambda})^{-1}_{ij}$$

$$+ \sum_{i=1}^{n} (\mathcal{F}_{B|B})^{-1}_{i0} \mu_i^T C^{-1} \mu_0 (\mathcal{F}_{A\setminus\Lambda})^{-1}_{i0}$$

$$+ (\mathcal{F}_{B|B})^{-1}_{00} \mu_0^T C^{-1} \mu_0 (\mathcal{F}_{A\setminus\Lambda})^{-1}_{00}.$$  \hspace{1cm} (30)

Notice that terms like $\mu_i^T C^{-1} \mu_j$ are just elements of the Fisher matrix and we can also use Equation (18) to express terms like $(\mathcal{F}_{B|B})^{-1}_{ij}$ in terms of determinants of minors of the Fisher matrix. Thus,

$$\text{Cov}(A^{ML}, B^{ML}) = \frac{1}{|\mathcal{F}_{B|B}| |\mathcal{F}_{A\setminus\Lambda}|}$$

$$\times \sum_{i,j=1}^{n} (-1)^{i+j+1} |\mathcal{F}_{B|B}| \mathcal{F}_0 (-1)^{i+j+1} |\mathcal{F}_{A\setminus\Lambda}|$$

$$+ \sum_{i=1}^{n} (-1)^{i+j+1} |\mathcal{F}_{B|B}| \mathcal{F}_0 |\mathcal{F}_{A\setminus\Lambda}|$$

$$+ |\mathcal{F}_{B|B}| \mathcal{F}_0 |\mathcal{F}_{A\setminus\Lambda}|.$$  \hspace{1cm} (31)

To simplify (32), first we combine term [1] and [2]:

$$[1] + [2] = \sum_{i=1}^{n} (-1)^{i+n+1} |\mathcal{F}_{B|B}|$$

$$\times \sum_{j=1}^{n} (-1)^{j+n+1} |\mathcal{F}_{A\setminus\Lambda}| + \mathcal{F}_0 |\mathcal{F}_{A\setminus\Lambda}|$$

$$= \sum_{j=1}^{n} (-1)^{i+n+1} |\mathcal{F}_{B|B}|$$

$$\times \sum_{j=1}^{n} (-1)^{j+n+1} |\mathcal{F}_{A\setminus\Lambda}| + \mathcal{F}_0 (-1)^{2(n+1)} |\mathcal{F}_{A\setminus\Lambda}|.$$  \hspace{1cm} (32)

Inserting (33) into (32), we have

$$\text{Cov}(A^{ML}, B^{ML}) = \frac{|\mathcal{F}_{A\setminus\Lambda}| |\mathcal{F}_{A\setminus\Lambda}|}{|\mathcal{F}_{B|B}| |\mathcal{F}_{A\setminus\Lambda}|}.$$  \hspace{1cm} (33)

Using Equations (28) and (29), the correlation between the best-fit values of parameters $A$ and $B$ estimated separately can be written as

$$\text{corr}(A^{ML}, B^{ML}) = \frac{\text{Cov}(A^{ML}, B^{ML})}{\sqrt{\text{Var}(A^{ML}) \times \text{Var}(B^{ML})}}$$

$$= \frac{|\mathcal{F}_{A\setminus\Lambda}| |\mathcal{F}_{A\setminus\Lambda}|}{|\mathcal{F}_{B|B}| |\mathcal{F}_{A\setminus\Lambda}|}.$$  \hspace{1cm} (34)

which is the same as Equation (22) up to a minus sign, thus completing our proof.

**ORCID iDs**

Y. Huang  \(https://orcid.org/0000-0003-4025-1536\)

G. E. Addison  \(https://orcid.org/0000-0002-2147-2248\)

C. L. Bennett  \(https://orcid.org/0000-0001-8839-7206\)

**References**

Abbott, T. M. C., Allan, S., Andersen, P., et al. 2019, ApJL, 872, L30

Addison, G. E., Huang, Y., Watts, D. J., et al. 2016, ApJ, 818, 132

Addison, G. E., Watts, D. J., Bennett, C. L., et al. 2018, ApJ, 853, 119

Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20

BICEP2 Collaboration, Keck Array Collaboration, Ade, P. A. R., et al. 2018, PhRvL, 121, 22130

Birrer, S., Treu, T., Rusu, C. E., et al. 2019, MNRAS, 484, 4726

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914

Box, G. E. P., & Cox, D. R. 1964, J. R. Stat. Soc. B, 26, 211

Calabrese, E., Slosar, A. c. v., Melchiorri, A., Smoot, G. F., & Zahn, O. 2008, PhRvD, 77, 123531

Gelman, A., Carlin, J., Stern, H., et al. 2013, Bayesian Data Analysis, Texts in Statistical Science (3rd edition; London/Boca Raton, FL: Chapman & Hall/CRC Press, Taylor & Francis)

Hannammeche, S., & Lewis, A. 2008, PhRvD, 77, 103013

Heavens, A. 2009, arXiv:0906.0664

Heavens, A., Fantaye, Y., Selentin, E., et al. 2017, PhRvL, 119, 101301

Henning, J. W., Sayre, J. T., Reichardt, C. L., et al. 2018, ApJ, 852, 97
