

χ^2 versus median statistics in supernova type Ia data analysisA. Barreira^{1,2,*} and P. P. Avelino^{2,3,†}¹*Centro de Física do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*²*Departamento de Física da Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*³*Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal*

(Received 20 July 2011; published 20 October 2011)

In this paper we compare the performances of the χ^2 and median likelihood analysis in the determination of cosmological constraints using type Ia supernovae data. We perform a statistical analysis using the 307 supernovae of the Union 2 compilation of the Supernova Cosmology Project and find that the χ^2 statistical analysis yields tighter cosmological constraints than the median statistic if only supernovae data is taken into account. We also show that when additional measurements from the cosmic microwave background and baryonic acoustic oscillations are considered, the combined cosmological constraints are not strongly dependent on whether one applies the χ^2 statistic or the median statistic to the supernovae data. This indicates that, when complementary information from other cosmological probes is taken into account, the performances of the χ^2 and median statistics are very similar, demonstrating the robustness of the statistical analysis.

DOI: [10.1103/PhysRevD.84.083521](https://doi.org/10.1103/PhysRevD.84.083521)

PACS numbers: 98.80.-k, 95.35.+d, 95.36.+x, 98.80.Es

I. INTRODUCTION

More than a decade ago, type Ia supernovae (SNIa) provided the first clear evidence in favor of cosmic acceleration [1,2]. Since then, the availability of ever larger, higher-quality SNIa data sets, as well as measurements using other cosmological probes, such as the cosmic microwave background (CMB) or the baryonic acoustic oscillations (BAO), have been providing overwhelming evidence for the existence of dark energy [3–5], a fluid with large negative pressure capable of driving the acceleration of the Universe.

In the so-called Λ CDM model, also known as the concordance model, the dark energy role is played by a cosmological constant Λ , responsible for approximately 73% of the energy density of the universe at the present day. The remaining percentage is mainly in the form of cold matter, most of which nonbaryonic and dark. Radiation is residual at the present time and the universe is spatially flat. Despite the good agreement with observational data, this model has little appeal on theoretical grounds since the value of Λ required to explain the observed cosmic acceleration is off by ~ 120 orders of magnitude from the standard quantum field theory prediction. The coincidence between our observing time and the time of the onset of cosmic acceleration is also puzzling [6,7]. Many other dark energy models have been proposed, with the most influential being the ones based on dynamical scalar fields. In these models, the energy density varies with time and suitable choices of the scalar field Lagrangian can relax some of the problems

associated with the cosmological constant (see, for instance, the review in [8,9]). As a result, an important step towards a better understanding of the dark energy involves further testing of its possible dynamical nature.

The determination of dark energy constraints from observational data requires a robust statistical data analysis. In the particular case of SNIa, the usual procedure is to carry out a standard χ^2 likelihood analysis. However, in [10,11] it has been argued that the median statistic could be a more reliable alternative. The median statistical analysis, despite having the drawback of being less constraining than the χ^2 analysis, has the strong advantage of requiring weaker assumptions about the data, thus yielding more trustworthy constraints. Moreover, the median is less vulnerable to the presence of “outliers,” which is also a significant advantage given the current uncertainties about the physics of SNIa.

In this work we revisit and extend the analysis of [11], updating it using the recent Union 2 SNIa compilation of the Supernova Cosmology Project (SCP) [4]. We compare the performances of χ^2 and median statistics in the determination of cosmological constraints using SNIa data, considering also CMB and BAO measurements. The layout of this paper is as follows. In Sec. II we describe the application of χ^2 and median statistics to SNIa data. We also discuss the use of additional information contained in the CMB and BAO. In Sec. III we present and discuss our results. Finally, we conclude in Sec. IV.

II. COSMOLOGICAL PROBES**A. SNIa**

Type Ia supernovae appear to be good “standard candles” and therefore they can serve as useful distance

*alex.mr.barreira@hotmail.com

†ppavelin@fc.up.pt

indicators. Constraints arise by comparing the theoretical distance modulus,

$$\mu_{\text{th}} = 5 \log_{10} \left(\frac{d_L(z)}{\text{Mpc}} \right) + 25, \quad (1)$$

with the observational distance modulus μ_{obs} inferred from the data, at the measured supernovae redshifts. In Eq. (1), $d_L(z)$ is the luminosity distance given by

$$d_L(z) = \frac{(1+z)c}{H_0 \sqrt{|\Omega_{k0}|}} S \left(\sqrt{|\Omega_{k0}|} \int_0^z \frac{dz'}{E(z')} \right), \quad (2)$$

where z is the cosmological redshift, H_0 is the present day value of the Hubble expansion rate and $E(z) = H(z)/H_0$. The function S is defined as

$$S(x) = \begin{cases} \sin x, & \Omega_{k0} < 0, \\ x, & \Omega_{k0} = 0, \\ \sinh x, & \Omega_{k0} > 0, \end{cases} \quad (3)$$

with the flat case ($\Omega_{k0} = 0$) being recovered taking the limit $\Omega_{k0} \rightarrow 0$ in Eq. (2). In this paper we consider

$$E(a) = \sqrt{\Omega_{m0} a^{-3} + \Omega_{de0} a^{-3(1+w)} + \Omega_{k0} a^{-2}}, \quad (4)$$

where $a = 1/(1+z)$ is the scale factor, Ω_{m0} and Ω_{de0} are, respectively, the present day values of the matter and dark energy fractional densities ($\Omega_i = \rho_i/\rho_c$, with ρ_c being the critical density) and $\Omega_{k0} = 1 - \Omega_{m0} - \Omega_{de0}$. The parameter w is the equation of state parameter of the dark energy (the ratio between the pressure and the energy density) which, for simplicity, we assume to be constant (see however [12,13]). If the dark energy role is played by a cosmological constant Λ then $\Omega_{de} = \Omega_\Lambda$ and $w = -1$.

1. χ^2 analysis

In the χ^2 statistical analysis, the likelihood P of the cosmological parameters Q is given by $P \propto \exp(-\chi^2/2)$, with

$$\chi^2 = \sum_{i=1}^N \left(\frac{\mu_{\text{obs}}(z_i) - \mu_{\text{th}}(z_i, Q)}{\sigma_i} \right)^2. \quad (5)$$

In Eq. (5), N is the number of supernovae in the data set and σ_i is the observational error associated with μ_{obs} at the redshift z_i . The χ^2 analysis assumes that: (i) the experimental results are statistically independent; (ii) there are no systematic errors; (iii) the statistical errors follow a Gaussian distribution; (iv) the standard

deviation of the statistical errors is equal to the observational uncertainty.

Presently, there is no strong evidence supporting that the supernovae magnitude errors are Gaussianly distributed and therefore the hypothesis (iii) and (iv) are quite strong. Moreover, the χ^2 statistic is highly susceptible to the presence of ‘‘outliers’’ in the data sets. This constitutes an extra concern, in particular, due to the uncertainties associated with the calibration of the SNIa light curves.

2. Median analysis

The fewer the assumptions one needs to make about the dataset, the more reliable the results derived from it are. The median statistical analysis discards the assumption of the Gaussianity of the errors, requiring only the use of hypothesis (i) and (ii).

Assuming that the experimental results are statistically independent and that there are no systematic errors present, one expects that after performing a sufficiently large number of measurements, approximately half of the values obtained will be above (or below) the correct mean value. In particular, if we perform N measurements, the probability that k of them will be above (or below) the median is given by the binomial distribution,

$$P_{\text{binomial}}(k, N) = \frac{2^{-N} N!}{k!(N-k)!}. \quad (6)$$

This way, given the data set $\mu_{\text{obs}}(z_i)$ [$i = 1 \dots N$], the likelihood of the cosmological parameters Q is obtained by counting the number of observational values that fall above (or below) the curve given by Eq. (1).

Despite being associated with less tighter constraints than the χ^2 statistic, the median statistic is not very sensitive to the presence of a few ‘‘outlier’’ SNIa objects. In [10], it has been shown how the presence of one or very few ‘‘ill’’ data points could severely distort a χ^2 analysis, while the median results remained approximately the same.

On the other hand, one should notice that the likelihood computed with the median statistic only accounts for the number of experimental points above or below the model’s curve, not differentiating between the various ways in which these points could be distributed. For instance, a model where the first half of the data points is above the model’s curve and the second half is below has the same likelihood (when computed with the median) as a model where the first point is above, the second below, the third above, the fourth below, and so forth. These two cases should not be indistinct since the first could turn out to be a terrible fit to the data. In [11] it has been shown how some modifications of the median statistic could alleviate this problem. In short, the modifications involved taking into account the size of the

largest continuous sequence found above (or below) the model's prediction, or the number of total continuous sequences obtained.

Here, we propose an alternative way to cope with this problem. Instead of counting the number of points that are above (or below) the model's prediction in the entire redshift range of the data set, one may divide the data set into redshift intervals with a certain number of SNIa objects and perform the counting in each interval. Suppose we divide the data set into n intervals, with N_j being the number of SNIa objects in the j -th interval ($j = 1 \dots n$). This way, the overall likelihood of the parameters Q is given by

$$P = \prod_{j=1}^n P_{\text{binomial}}(k_j, N_j), \quad (7)$$

where k_j is the number of points that, in the j -th interval, are above (or below) the theoretical curve given by the parameters. This way, by properly dividing the data set into groups of supernovae, we are more likely to avoid pathological situations in which very large sequences of SNIa above (or below) the median value are present.

B. The CMB shift parameter

The CMB shift parameter R is defined by

$$\begin{aligned} R_{\text{th}} &\equiv \sqrt{\Omega_{m0}} \frac{H_0}{c} \frac{d_L(z_{\text{dec}})}{(1+z_{\text{dec}})} \\ &= \frac{\sqrt{\Omega_{m0}}}{\sqrt{|\Omega_{k0}|}} S\left(\sqrt{|\Omega_{k0}|} \int_0^{z_{\text{dec}}} \frac{dz'}{E(z')}\right), \end{aligned} \quad (8)$$

with the redshift of decoupling $z_{\text{dec}} \approx 1090.97$ [5,14]. Following the WMAP 7-year results [5] we take $R_{\text{obs}} = 1.725$ with an error $\sigma_R = 0.018$. The likelihood derived from the shift parameter is then $P_{\text{CMB}} \propto \exp(-\chi_{\text{CMB}}^2/2)$, with

$$\chi_{\text{CMB}}^2 = \left(\frac{R_{\text{th}} - R_{\text{obs}}}{\sigma_R}\right)^2. \quad (9)$$

In the case of the CMB, there are spatial correlations between the data points which do not satisfy hypothesis (i), crucial for the use of the median statistic. However, it turns out that assuming that R is Gaussianly distributed around R_{obs} with standard deviation σ_R provides an efficient summary of the information encoded in the full CMB data [15,16].

C. The BAO scale

The baryonic acoustic oscillations imprinted in the CMB manifest themselves today in the large-scale distribution of galaxies. The BAO signature on large scales was found in [17], when a small ‘‘bump’’ in the two-point correlation

function of red-luminous galaxies was measured. Cosmological constraints arise via the position of the ‘‘bump’’, which is related to the quantity

$$\begin{aligned} A_{\text{th}}(z) &= \sqrt{\Omega_{m0}} E(z)^{-(1/3)} \\ &\times \left[\frac{1}{z\sqrt{|\Omega_{k0}|}} S\left(\sqrt{|\Omega_{k0}|} \int_0^z \frac{dz'}{E(z')}\right) \right]^{2/3}. \end{aligned} \quad (10)$$

The likelihood is $P_A \propto \exp(-\chi_A^2/2)$, with

$$\chi_A^2 = \left(\frac{A_{\text{th}}(z_{\text{BAO}}) - A_{\text{obs}}}{\sigma_A}\right)^2, \quad (11)$$

where $z_{\text{BAO}} = 0.35$, $A_{\text{obs}} = 0.469$ and $\sigma_A = 0.017$ [18,19]. Just like in the case of the CMB, there are spatial correlations between the BAO data points which prevent the use of the median statistic. However, it is usually safe to assume that A is Gaussianly distributed [17,18].

III. RESULTS

A. SNIa constraints

Following the discussion in Sec. II A, we divide the SNIa Union 2 dataset [4], which contains a total of $N = 307$ SNIa objects, in groups of approximately 75 SNIa each. The supernovae are ordered by increasing redshift with the first three groups having 76 supernovae and the fourth group 79. Several other ways to divide the set are possible. However, this choice is not critical for our conclusions.

Figures 1 and 2 show the 68.3%, 95.4%, and 99.7% confidence level contours on Ω_{m0} and $\Omega_{\Lambda0}$, obtained from the Union 2 SNIa data set using the χ^2 and median statistical analysis, respectively. Figures 3 and 4 are similar to Figs. 1 and 2, except that now the SNIa constraints are

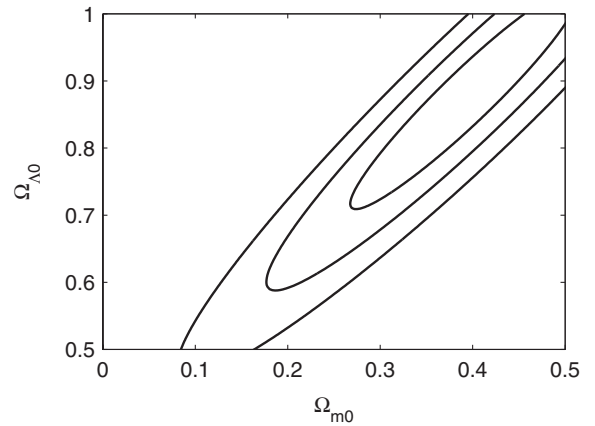


FIG. 1. 68.3%, 95.4% and 99.7% confidence level contours on Ω_{m0} and $\Omega_{\Lambda0}$, obtained from the Union 2 SNIa data set using the χ^2 statistical analysis.

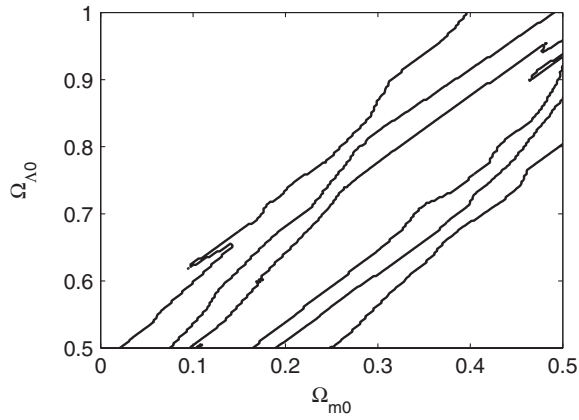


FIG. 2. Same as Fig. 1 but using the median statistical analysis.

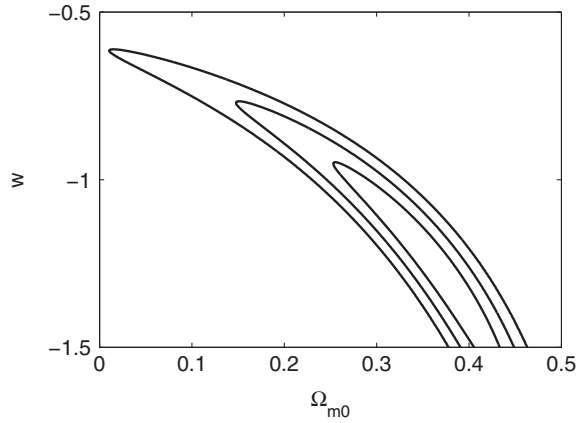


FIG. 3. 68.3%, 95.4% and 99.7% confidence level contours on Ω_{m0} and w (assuming that $\Omega_{k0} = 0$), obtained from the Union 2 SNIa data set using the χ^2 statistical analysis.

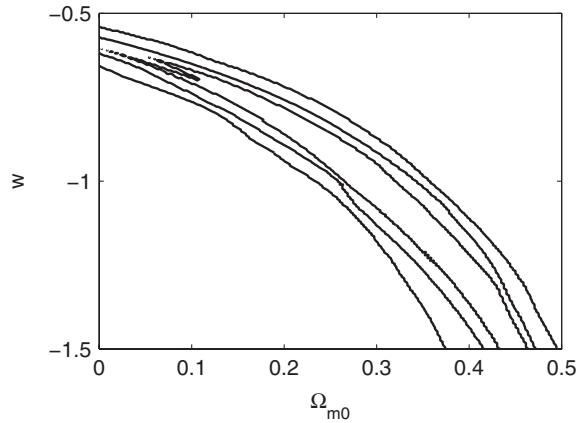


FIG. 4. Same as Fig. 3 but using the median statistical analysis.

on the (Ω_{m0}, w) plane and the condition $\Omega_{k0} = 0$ was assumed. In both cases the Hubble parameter today was held fixed at $H_0 = 70.2$ km/s/Mpc [5]. As expected, the constraints using the χ^2 statistic are tighter, consequence of the stronger assumptions it makes about the data. Nevertheless, the constraints obtained using the median statistical analysis are not as bad as one could originally fear. The median contours are more “stretched” than the χ^2 ones, but their “width” is similar. The median statistical constraints are in principle more reliable, since they do not assume Gaussianity of the SNIa magnitude error distribution.

B. SNIa + CMB + BAO constraints

Figures 5 and 6 show the 68.3%, 95.4% and 99.7% confidence level contours on Ω_{m0} and $\Omega_{\Lambda0}$, obtained from the Union 2 SNIa data set (using the χ^2 and median

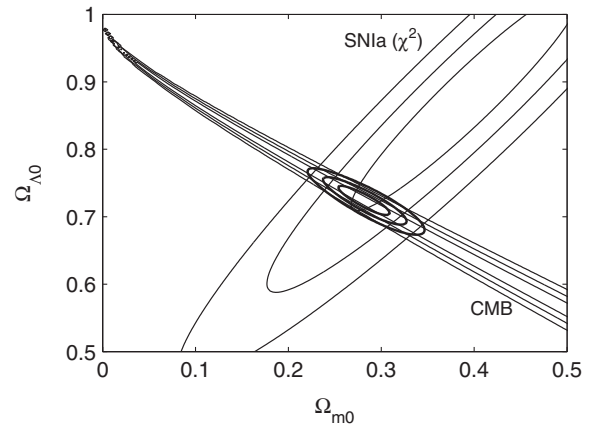


FIG. 5. 68.3%, 95.4% and 99.7% confidence level contours on Ω_{m0} and $\Omega_{\Lambda0}$, obtained from the Union 2 SNIa data set, combined with constraints on the CMB shift parameter R , using the χ^2 statistical analysis.

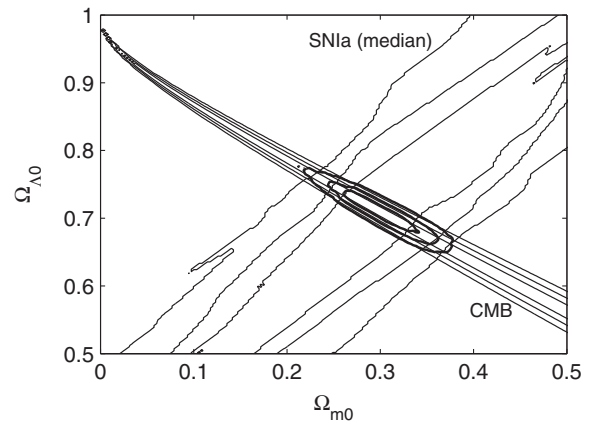


FIG. 6. Same as Fig. 5 but using the median statistical analysis for the SNIa data.

TABLE I. Constraints on Ω_{m0} , Ω_{de0} and w , obtained from SNIa (χ^2 and median analysis), CMB and BAO data, as well as their combinations. The constraints are at the 1σ level.

Fit	Ω_{m0}	Ω_{k0}	w
SNIa(χ^2) + CMB	[0.254; 0.310]	[-0.043; 0.056]	-1(fixed)
SNIa(median) + CMB	[0.260; 0.342]	[-0.061; 0.083]	-1(fixed)
SNIa(2) + CMB + BAO	[0.269; 0.313]	[-0.031; 0.049]	-1(fixed)
SNIa(median) + CMB + BAO	[0.275; 0.333]	[-0.042; 0.064]	-1(fixed)
SNIa(2) + CMB + BAO	[0.264; 0.308]	0(fixed)	[-1.084; -0.955]
SNIa(median) + CMB + BAO	[0.266; 0.312]	0(fixed)	[-1.058; -0.902]

statistical analysis, respectively), combined with constraints on the CMB shift parameter R . The differences between the results obtained with the χ^2 and median statistics are now significantly reduced. This is related to the fact that, as shown in Figs. 1 and 2, the “width” of the

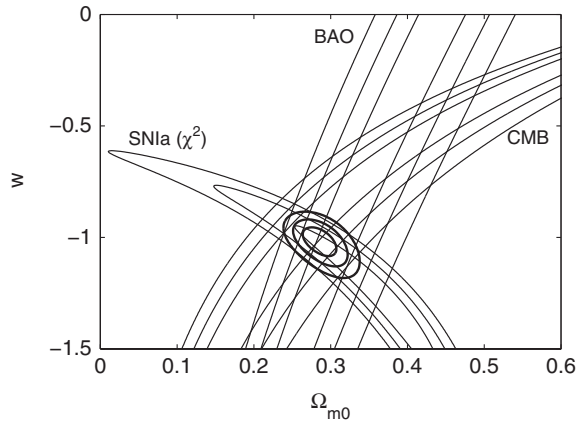


FIG. 7. 68.3%, 95.4% and 99.7% confidence level contours on Ω_{m0} and w (assuming that $\Omega_{k0} = 0$), obtained from the Union 2 SNIa data set, combined with constraints on the CMB shift parameter R and the BAO scale, using the χ^2 statistical analysis.

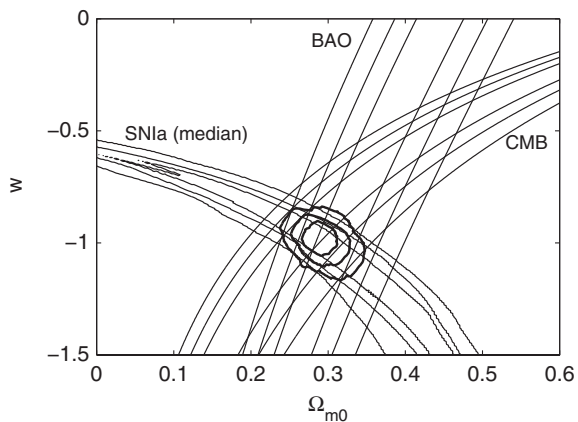


FIG. 8. The same as Fig. 7 but using the median statistical analysis for the SNIa data.

SNIa contours, obtained using the χ^2 and median analysis, is very similar. The CMB contours “cross” the SNIa contours, rendering similar combined constraints in both cases. Considering the BAO data (not shown) tightens the constraints slightly (see Table I).

Figures 7 and 8 show the combined constraints on Ω_{m0} and w (assuming that $\Omega_{k0} = 0$), obtained from SNIa (using χ^2 and median analysis, respectively), CMB and BAO data. Again, we see that the combined constraints are weakly dependent on whether one derives the constraints from SNIa with the χ^2 or with the median statistical analysis.

Table I presents a summary of the constraints, at the 1σ level, obtained from different data combinations and model assumptions. It shows that the median is almost as constraining as the χ^2 statistical analysis, if combined with additional CMB and BAO constraints. Hence, the strong assumption that the SNIa measurements are Gaussianly distributed is not necessary in order to obtain tight constraints. On the other hand, since the χ^2 and median statistical analysis yield similar cosmological constraints when the SNIa data is combined with other cosmological probes, our results may also be interpreted as providing additional validation of the results obtained using the standard χ^2 statistical analysis.

IV. CONCLUSION

In this paper we compared the cosmological constraints derived from the Union 2 SNIa dataset using χ^2 and median statistics. In the absence of CMB and BAO constraints, we have shown that the χ^2 statistic yields tighter cosmological constraints than the median statistic, as a result of the stronger assumptions it makes about the SNIa magnitude error distribution. On the other hand, when CMB and BAO information is taken into account, the performances of both statistics are very similar. Hence, we conclude that the assumption of the Gaussian distribution of the errors in the SNIa analysis does not appear to be critical in the determination of cosmological constraints, provided that complementary information from other cosmological probes is also taken into account.

- [1] A. G. Riess *et al.* (Supernova Search Team), *Astron. J.* **116**, 1009 (1998).
- [2] S. Perlmutter *et al.* (Supernova Cosmology Project), *Astrophys. J.* **517**, 565 (1999).
- [3] J. Frieman, M. Turner, and D. Huterer, *Annu. Rev. Astron. Astrophys.* **46**, 385 (2008).
- [4] R. Amanullah *et al.*, *Astrophys. J.* **716**, 712 (2010).
- [5] E. Komatsu *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **192**, 18 (2011).
- [6] P. P. Avelino, *Phys. Lett. B* **611**, 15 (2005).
- [7] A. Barreira and P. P. Avelino, *Phys. Rev. D* **83**, 103001 (2011).
- [8] E. J. Copeland, M. Sami, and S. Tsujikawa, *Int. J. Mod. Phys. D* **15**, 1753 (2006).
- [9] M. Li, X.-D. Li, S. Wang, and Y. Wang, *Commun. Theor. Phys.* **56**, 525 (2011).
- [10] J. R. Gott, III, M. S. Vogeley, S. Podariu, and B. Ratra, *Astrophys. J.* **549**, 1 (2001).
- [11] P. P. Avelino, C. J. A. P. Martins, and P. Pinto, *Astrophys. J.* **575**, 989 (2002).
- [12] P. P. Avelino, A. M. M. Trindade, and P. T. P. Viana, *Phys. Rev. D* **80**, 067302 (2009).
- [13] P. P. Avelino, L. Losano, and J. J. Rodrigues, *Phys. Lett. B* **699**, 10 (2011).
- [14] W. Hu and N. Sugiyama, *Astrophys. J.* **471**, 542 (1996).
- [15] Y. Wang and P. Mukherjee, *Phys. Rev. D* **76**, 103533 (2007).
- [16] Ø. Elgarøy and T. Multamäki, *Astron. Astrophys.* **471**, 65 (2007).
- [17] D. J. Eisenstein *et al.* (SDSS Collaboration), *Astrophys. J.* **633**, 560 (2005).
- [18] W. J. Percival *et al.* (SDSS Collaboration), *Mon. Not. R. Astron. Soc.* **401**, 2148 (2010).
- [19] B. A. Reid *et al.*, *Mon. Not. R. Astron. Soc.* **404**, 60 (2010).