


Nonparametric analysis of the Hubble diagram with neural networks

Lorenzo Giambagli^{1,2,3} , Duccio Fanelli^{1,3}, Guido Risaliti^{1,4}, and Matilde Signorini^{1,4}

¹ Dipartimento di Fisica e Astronomia, Università di Firenze, Via G. Sansone 1, 50019 Sesto Fiorentino, Firenze, Italy
e-mail: lorenzo.giambagli@unifi.it

² naXys – Namur Center for Complex Systems, University of Namur, Rue Grafé 2, 5000 Namur, Belgium

³ INFN and CSDC, Via Sansone 1, 50019 Sesto Fiorentino, Firenze, Italy

⁴ INAF – Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy

Received 24 February 2023 / Accepted 10 July 2023

ABSTRACT

The recent extension of the Hubble diagram of supernovae and quasars to redshifts much higher than 1 prompted a revived interest in nonparametric approaches to test cosmological models and to measure the expansion rate of the Universe. In particular, it is of great interest to infer model-independent constraints on the possible evolution of the dark energy component. Here we present a new method, based on neural network regression, to analyze the Hubble diagram in a completely nonparametric, model-independent fashion. We first validated the method through simulated samples with the same redshift distribution as the real ones, and we discuss the limitations related to the “inversion problem” for the distance-redshift relation. We then applied this new technique to the analysis of the Hubble diagram of supernovae and quasars. We confirm that the data up to $z \sim 1-1.5$ are in agreement with a flat Λ cold dark matter model with $\Omega_M \sim 0.3$, while ~ 5 -sigma deviations emerge at higher redshifts. A flat Λ cold dark matter model would still be compatible with the data with $\Omega_M > 0.4$. Allowing for a generic evolution of the dark energy component, we find solutions that suggest an increasing value of Ω_M with redshift, as predicted by interacting dark sector models.

Key words. methods: data analysis – methods: analytical – cosmological parameters

1. Introduction

The Hubble diagram (i.e., the distance-redshift relation) describes the expansion of the Universe with time and is one of the fundamental tools of observational cosmology. The “kinematic” information encoded in this diagram includes the Hubble parameter H_0 (from the first-order derivative at redshift $z = 0$) and the acceleration parameter (from the second-order derivative). When a dynamical model is adopted, its physical parameters can be derived from the fit of the Hubble diagram. Typical examples are the estimate of the matter density at $z = 0$, Ω_M , within a flat Λ cold dark matter (CDM) model, or the evaluation of Ω_M and Ω_Λ within a non-flat Λ CDM model. Moreover, the physical meaning of the relevant parameters to some extent reflects the chosen model. Likewise, the obtained numerical estimates are also model-dependent: if we assume, for example, that data follow a Λ CDM model, with prescribed Ω_M and nonzero curvature, it is easy to demonstrate through numerical simulations that, if a flat Λ CDM is adopted, the best-fit value of Ω_M will be different from the correct (simulated) one.

In the past few years, possible new physics beyond the flat Λ CDM model has been suggested by several observational results, such as the mismatch between the direct measurements of H_0 in the local Universe (Riess et al. 2019; Wong et al. 2019) and the extrapolations based on the cosmic microwave background, the comparison between the high- and low- multipole spectra of the cosmic microwave background (Di Valentino et al. 2021), and the tension between the power spectrum of density perturbations measured on different scales (Macaulay et al. 2013; Battye et al. 2015; Lin & Ishak 2017; Heymans et al. 2021;

Nunes & Vagnozzi 2021). Recently, a significant deviation from the flat Λ CDM model has been observed in the Hubble diagram at high redshifts, an area populated with quasars and γ -ray bursts: while no significant tension is found at $z < 1.5$ with supernovae, quasars, or γ -ray bursts, the data at $z > 1.5$ suggest a slower expansion of the Universe than predicted by the flat Λ CDM model (Risaliti & Lusso 2019; Demianski et al. 2017; Lusso et al. 2019, 2020). These results make it particularly important to analyze the Hubble diagram in a way that is as model-agnostic as possible, in order to obtain an “absolute scale” for the comparison with specific models, and to infer the global, “cosmographic” properties of the expansion, which, in turn, could suggest the optimal class of model to fit to the data.

Cosmographic expansions (Aviles et al. 2014; Capozziello et al. 2020; Bargiacchi et al. 2021) represent a viable approach to achieving this goal. The method is based on a standard fitting procedure and assumes that observational data can be interpolated by an appropriate series of functions, truncated to include a limited number of terms (and hence a limited number of free parameters). While this is not dependent on a specific physical model, it still relies on the flexibility of the chosen functions to reproduce the shape of the observational Hubble diagram. Moreover, cosmographic techniques are rigorously valid only within a convergence radius, which is $z = 1$ for standard methods (Cattoën & Visser 2007). At higher redshifts, no method has an absolute validity based on mathematical principles, and the effectiveness of the cosmographic analysis relies on the similarity between the chosen expansion functions and the actual shape of Hubble diagram.

An example of a robust, well-checked, nonparametric approach is that based on Gaussian process (GP) regression (Holsclaw et al. 2010; Seikel et al. 2012; Shafieloo et al. 2012), which has been used to test the hypothesis of a constant density of the dark energy term (i.e., the cosmological constant Λ). Despite their flexibility, however, GPs may underestimate the error associated with predictions (Colgáin 2021) and come with an intrinsic convergence problem for $z > 1$.

Starting from these premises, we here propose, and consequently apply, a novel analysis framework for the Hubble diagram based on neural network (NN) regression (see Dialektopoulos et al. 2022 for a similar approach). Deep feed-forward fully connected NNs are well-known universal approximators. Their ability to represent functions extends far beyond what is required for the particular problem at hand. However, the core of their efficacy lies in the assumption that the proper regression function results from a collection of several multilevel hierarchical factors (or features) that could enable one to account for unknown features in the analysis that bear – at least some – cosmological relevance. To sum up, we try to merge the concept of features with cascading relevance appropriate for cosmographic expansion with the need for convergence and the high-function representation capabilities typical of kernel methods or GPs.

We first describe the method and check its reliability with simulated data sets. Then we apply it to a Hubble diagram at high redshifts, showing a high-redshift inconsistency with the Λ CDM model. Finally, we speculate on the class of model that could fix the discrepancy.

2. The cosmological background

In a Friedmann–Robertson–Walker Universe, the luminosity distance of an astrophysical source is related to the redshift through the equation

$$d_L = \frac{c(1+z)}{H_0 \sqrt{-\Omega_K}} \sin \left(\sqrt{-\Omega_K} \int_0^z dz' \frac{H_0}{H(z')} \right), \quad (1)$$

where $H(z)$ is the Hubble function and Ω_K stands for the curvature parameter, defined as $\Omega_K = 1 - \sum_i \Omega_i$, with Ω_i representing the density of the constituents of the Universe, normalized to the closure density. In the simplest form – assuming a flat Universe, a constant total content of matter in the Universe, and a cosmological constant and considering the redshift range where standard candles are observed (i.e., $z < 7$, where the contribution of the radiation and neutrino terms is negligible) – $H(z) = H_0 \sqrt{\Omega_M (1+z)^3 + 1 - \Omega_M}$. However, a wide range of different physical and cosmological models have been considered in the literature, including a nonzero curvature, an evolving dark energy density, and/or interactions between dark energy and dark matter. In this work, we analyze a subset of these models, represented by the equation

$$H(z) = H_0 \sqrt{\Omega_M (1+z)^3 + (1 - \Omega_M) e^{3 \int_0^z \frac{1+w(z')}{1+z'} dz'}}, \quad (2)$$

where $w(z)$ is a generic redshift evolution of the dark energy component density. Our main goal is to test the consistency of the flat Λ CDM hypothesis (which amounts to setting $w = -1$ in Eq. (2)) with the present Hubble diagram of supernovae and quasars, and draw a comparison with other possible functional forms for $w(z)$, as proposed in the literature. To this aim, we carried out a nonparametric fit via a suitably designed NN. This enabled us to reach conclusions on the predicted profile of $w(z)$ without relying on any a priori assumption.

One key problem in any nonparametric reconstruction attempt is the so-called inversion problem: it is easy to demonstrate that the inversion of Eq. (2), which involves the first and second derivatives of $H(z)$ (see, e.g., Seikel et al. 2012), is inherently unstable, due to a strong dependence on the Ω_M and H_0 parameters; in particular, a change in the quantity $H_0^2 \Omega_M$ by as little as 0.1% can alter the predicted value of $w(z)$ by orders of magnitude and/or flip its sign. As a consequence, constraints on $w(z)$ at very low redshifts can be obtained, but the uncertainties become very large beginning at $z \sim 0.5$. This makes it difficult to reach hard conclusions about the supposed consistency of available data with the reference scenario with $w = -1$. In principle, better data could help reduce the uncertainties. While we will discuss this issue in more detail in a dedicated paper, here we just mention the relevant point for the present work: it is not possible to obtain significant information on $w(z)$ from the Hubble diagram without (a) assuming some analytic form of the function and/or (b) having a combined estimate of Ω_M and H_0 with a much higher precision than what is available today and in the foreseeable future. There are only two possible direct ways to overcome this limitation: either we restrict our analysis to very narrow ranges of the parameters, or we constrain the shape of the function $w(z)$. Since neither of these approaches is satisfactory (and both have already been explored in the literature), we chose a different strategy. We did not attempt to carry out a full inversion of Eq. (2). On the contrary, we overcame the aforementioned numerical problems by estimating the quantity

$$I(z) = \int_0^z \frac{w(z') + 1}{1+z'} dz', \quad (3)$$

which can be determined from the observational data by solely invoking the first derivative of $H(z)$ (see Cárdenas 2015 for an early application of this technique). We notice that within the Λ CDM model, $w = -1$ implies $I(z) = 0$. As an obvious limitation, we will just recover the integral of the physical quantity of interest, the function $w(z)$: the degeneracy on $w(z)$ implies that different forms of $w(z)$ lead to indistinguishable shapes of $I(z)$. Nonetheless, we can achieve some remarkable results. First, we can compare the results on $I(z)$ with the prediction of the flat Λ CDM model: an inconsistency here would be powerful and general proof of a tension between the model and the data (note that the opposite is not true: an agreement based on the analysis of $I(z)$ does not necessarily imply an invalidation of the Λ CDM model). More generally, we can explore the family of $w(z)$ functions that lead to the observation-based reconstruction of $I(z)$ to determine which class of physical model can reproduce the observed Hubble diagram.

3. Regression via deep neural networks

For our purposes we have chosen to deal with a fully connected feed-forward architecture, as illustrated in the supplementary information (SI) Appendix B. Function (3) is hence approximated by a suitable NN, denoted as I_{NN} , to be determined via the following apposite optimization procedure. After a few manipulations, as detailed in the SI, the data set takes the form $\mathcal{D} = \{(z^{(i)}, y^{(i)}, \Delta y^{(i)})\}$ with $i \in 1 \dots |\mathcal{D}|$, where $y^{(i)}$ is connected to the modulus of luminosity distance, $d_L^{(i)}$, and $\Delta y^{(i)}$ stands for the associated empirical error. The predictions, $y_{\text{pred}}^{(i)}$, and the supplied input, $y^{(i)}$, are linked via

$$y_{\text{pred}}^{(i)} = \int_0^{z^{(i)}} dz' \left[\Omega_M (1+z')^3 + (1 - \Omega_M) e^{3 I_{NN}(z')} \right]^{-\frac{1}{2}}. \quad (4)$$

We note that the prediction is a functional of I_{NN} , the NN approximation that constitutes the target of the analysis. To carry out the optimization, we introduced the loss function

$$L(I_{\text{NN}}, \mathcal{D}) = \sum_{i=1}^{|\mathcal{D}|} \left(\frac{y^{(i)} - y_{\text{pred}}^{(i)}}{\Delta y^{(i)}} \right)^2.$$

The weights of the network that ultimately defines I_{NN} are tuned so as to minimize the above loss function via conventional stochastic gradient descent (SGD) methods. The hyperparameters were optimized with mock data samples, as illustrated in the SI. To quantify the statistical errors Δy_{pred} (associated with the predictions) and ΔI_{NN} (referring to the approximating NN), we implemented a bootstrap procedure, which is detailed in the SI. The code is freely available online¹.

This regression scheme was tested against a selection of mock data samples. In carrying out the test we considered:

(A) A sample of 4000 sources with no dispersion, with a flat distribution in $\log(z)$ between $z = 0.01$ and $z = 6$, and following a flat Λ CDM model with $\Omega_{\text{M}} = 0.3$ and $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.7$. This sample (as well as the next on the list) represents a highly idealized, hence unrealistic, setting. It is solely used as a reference benchmark model for preliminary consistency checks.

(B) The same as above, but the model used is a Chevallier–Polarski–Linder (CPL) parametrization, which assumes a dark energy equation of state that varies with the redshift as $w(z) = w_0 + w_a \frac{z}{1+z}$ (Chevallier & Polarski 2001), with $w_0 = -1.5$ and $w_a = 0.5$. We note that this choice of the parameters would be hard to justify from a physical point of view. In particular, it implies a turning point in the Hubble parameter $H(z)$, which violates the null-energy condition. However, even if this scenario can be considered unphysical, it serves our purpose of testing our regression method with extreme models.

(C) A sample with the same size, redshift distribution, and dispersion as the Pantheon supernova Ia sample (Scolnic et al. 2018), assuming a flat Λ CDM model with $\Omega_{\text{M}} = 0.3$.

(D) A Pantheon-like sample, as above, but assuming a CPL model with $w_0 = -1.5$ and $w_a = 0.5$.

(E) A sample with the same size and redshift distribution as the combined Pantheon (Scolnic et al. 2018) and quasar (Lusso et al. 2020) samples. The quasar sample consists of 2244 sources with redshifts in the $z = 0.5$ – 7.5 range (the whole Lusso et al. 2020 sample contains 2421 objects, but the 178 ones at redshift $z < 0.5$ are not considered in this analysis). We assumed the same dispersion as in the real sample and a flat Λ CDM model with $\Omega_{\text{M}} = 0.3$.

(F) The same as setting (E) but assuming a CPL model with $w_0 = -1.5$ and $w_a = 0.5$.

More specifically, we generated synthetic data following these different recipes. The regression scheme, as implemented via the NN, enables us to solve an inverse problem, from the data back to the underlying physical model. The correspondence between postulated and reconstructed physical instances readily translates to a reliable metric for gauging the performance of the proposed procedure, in a fully controllable environment and prior to being applied to the experimental data set. The analysis of settings A and B is discussed in the SI and confirms that our NN method can consistently recover the “true” model and parameters with simulated data of (unrealistic) high quality.

The outcome of the analysis for settings C (top left), D (top right), E (bottom left), and F (bottom right) is displayed in Fig. 1. Both $I_{\text{NN}}(z)$ (the NN approximation for $I(z)$) and $y_{\text{pred}}(z)$ are represented as a function of the redshift, z . For settings E and F,

the associated mean loss is also plotted against the parameter Ω_{M} , which can be freely modulated to explore different scenarios. Working with a data set of type C cannot yield definite conclusions: the NN is unable to recover the correct value of Ω_{M} because different Λ CDM models ($I_{\text{NN}}(z) \simeq 0$, within the explored range) provide an equally accurate interpolation of the (simulated) data within statistical errors. The above degeneracy is, however, removed when extending the examined sample so as to include quasars (see the bottom-left panel of Fig. 1, which shows data set E). In this case, the minimum displayed by the loss function points to $\Omega_{\text{M}} = 0.3$, the value assumed in the simulations; the corresponding function $I_{\text{NN}}(z)$ is approximately equal to zero (green shadowed domain) within the errors, which is at variance with what is found by employing the other chosen values of Ω_{M} . Data sets D and F (rightmost panels in Fig. 1) return similar conclusions when operating with data generated according to a CPL prescription. Working with supernovae (over a limited range in z) does not allow us to distinguish between the Λ CDM and CPL model, while the underlying model, assumed for data generation, is correctly singled out when quasars are accounted for (green shadowed region that encloses the dashed line, which represents the exact profile), that is, when extending the data set to higher redshifts. Overall, our results from working on synthetic data suggest that (a) the regression method is reliable and (b) with the current Hubble diagram of supernovae, it is not possible to test the Λ CDM model against possible extensions, such as the CPL model with “phantom-like” dark energy. Such a degeneracy is removed with a combined supernova+quasar sample extending up to $z \sim 7$.

Motivated by the outcome of these simulations, we applied the NN to the experimental data set consisting of the Pantheon supernova sample and the Lusso et al. (2020) quasar sample. The quasar luminosity distances and their errors are estimated from the observational data (X-ray and UV fluxes) following the procedure described in Lusso et al. (2020). In summary, the X-ray and UV data are first fitted in narrow redshift bins in order to derive a cosmology-independent slope, α , of the X-ray to UV relation. This value and its uncertainty are used to derive luminosity distances on an arbitrary scale. The absolute calibration, β , was obtained from the cross-match of the quasar and supernova sample in the common redshift interval. This calibration has a negligible uncertainty with respect to the other components on the error on the luminosity distances: the flux measurement errors, the intrinsic dispersion of the relation, and the error on the slope. In principle, rather than fitting the so-derived luminosity distances, it would be preferable to use the observables (i.e., the fluxes) and to marginalize over the parameters α and β . In practice, the two methods provide identical results, and the use of luminosity distances and Eq. (4) makes the analysis much faster and easier to implement within a NN method.

The results of the NN-based fits are shown in Fig. 2. The shape of $I(z)$ is clearly not consistent with the flat Λ CDM model ($I(z) \equiv 0$). This is the main result of our work, which we arrived at without assuming any a priori knowledge on the function $I(z)$.

As a next step, we introduced a dedicated indicator to quantitatively measure the compatibility of the examined data with the reference Λ CDM model. We naively accessed the distance of the fitted profile, y_{pred} , to the reference $y_{\Lambda\text{CDM}}$ ($I = 0$) curve and divided it by the error associated with the fitted function Δy_{pred} . Assuming that the computed ratio (averaged over z) is smaller than the unit, the distance between y_{pred} and $y_{\Lambda\text{CDM}}$ is eclipsed by statistical uncertainty and thus Λ CDM cannot be ruled out as a candidate explanatory model. The above procedure has been verified (see the SI) and yields a scalar indicator that fulfills the

¹ <https://github.com/Jamba15/Cosmological-Regression-with-NN.git>

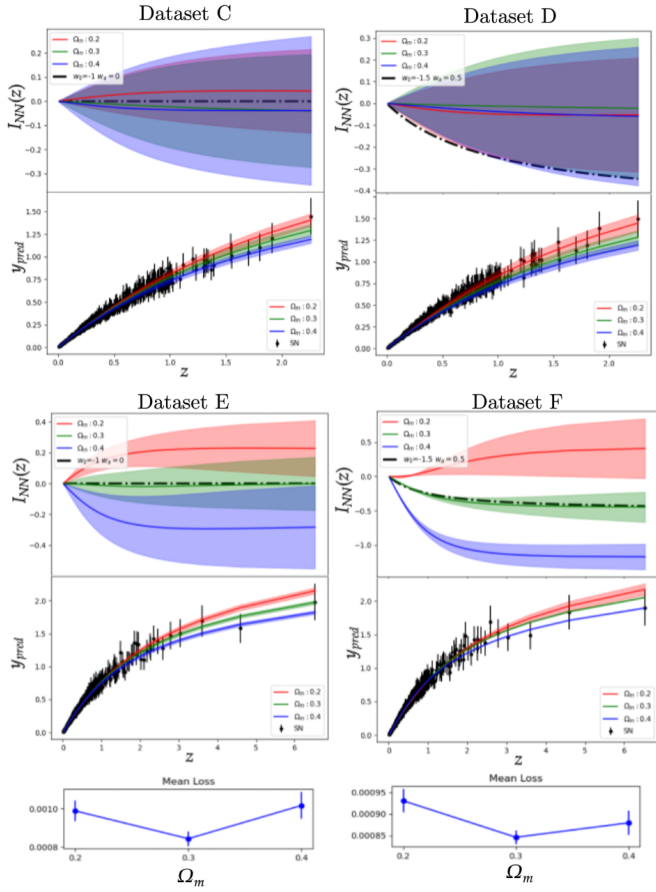


Fig. 1. $I(z)$ Results of the NN analysis of the Hubble diagram of simulated data. Top left: data set C, with the same redshift distribution and dispersion as the Pantheon supernova sample. Bottom left: data set E, where combined Pantheon and quasars are considered. In this case, the NN is able to identify the model assumed for data generation (the green shadowed region contains the exact profile for $I_{NN}(z)$, depicted with a dashed line). The corresponding loss function is also shown and displays a minimum at the correct value of Ω_M . Top right: Pantheon-like sample assumed for a CPL generative model (data set E). The NN is unable to distinguish between different scenarios (Λ CDM vs. CPL). Bottom right: CPL model with the inclusion of quasars. The degeneracy is resolved, and the NN can correctly identify the underlying model (see the dashed line). The loss shows a minimum for the correct value of Ω_M , which yields the green shadowed solution for $I_{NN}(z)$ vs. z .

purpose of quantifying the sought distance, normalized to the associated error. This is denoted by $\Delta_{\Lambda\text{CDM}}$ and takes the form

$$\Delta_{\Lambda\text{CDM}}(\mathcal{D}, I_{NN}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} \frac{\delta y_{\text{pred}}^{\Lambda\text{CDM}}(I_{NN}; z^{(i)})}{\Delta y_{\text{pred}}(I_{NN}; z^{(i)})}. \quad (5)$$

The fitted integral function I_{NN} is deemed compatible with the Λ CDM model if $\Delta_{\Lambda\text{CDM}} < 1$. When this condition holds true, the predictions deviate from Λ CDM by an amount that, on average, is smaller than the corresponding prediction error. The indicator in Eq. (5) was computed for different mock samples, mimicking Λ CDM, with progressively increasing errors sizes, Δy . This error size is assumed uniform across data points and varied from zero to 0.15, thus including the value that is believed to apply to real data: ~ 0.14 . This information is used as a reference benchmark to interpret the results of the analysis for the Pantheon + quasar experimental data set. To sum up our conclusions (see the SI), the portion of the data set at low redshifts is compatible

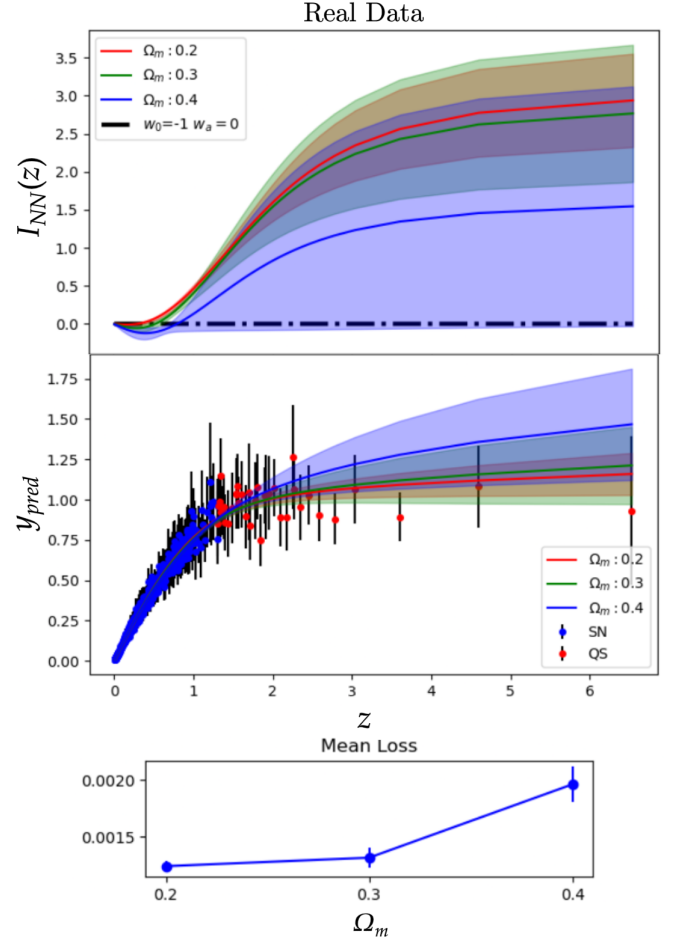


Fig. 2. Results of the NN analysis of the Hubble diagram of supernovae (blue points in the middle panel) and quasars (red points). Top panel: estimated values of $I(z)$ for different values of Ω_M . Central panel: Hubble diagram with the reconstructed best-fit function obtained from the NN analysis. Bottom panel: Loss values for different values of Ω_M . Notice that the solution visually closer (accounting for statistical errors) to the reference Λ CDM profile yields a significantly larger loss value and, as such, should be disregarded. The Loss is indeed nearly flat for $\Omega_M < 0.3$.

with a Λ CDM model with $\Omega_M = 0.3$, within statistical errors. Conversely, for $z > 2$ (notably quasars), $\Delta_{\Lambda\text{CDM}}$, as computed after available experiments, is 5σ from the expected mean value. Hence, accounting for quasars enables us to conclude that the Λ CDM model is indeed extremely unlikely.

Figure 3 depicts the results where the best-fit $I(z)$ for $\Omega_M = 0.3$ (the same as in the upper panel of Fig. 2) is plotted in logarithmic scale and compared to $I_{\text{MATTER}}(z) = \log(z)$, the function obtained from Eq. (3) by assuming $w(z) \equiv 0$ (i.e., a pure matter contribution). We recall that a cosmological constant, or equivalently a dark energy component with constant energy, implies $w(z) \equiv -1$ and $I(z) \equiv 0$. It is therefore tempting to speculate as follows when qualitatively analyzing the profile of $I(z)$: the redshift intervals with negative derivative represent a dark energy component with density increasing over time (the “phantom” dark energy scenario); the intervals with positive derivatives, smaller than the constant derivative of $I_{\text{MATTER}}(z)$, represent a dark energy component with decreasing density; last, the intervals where the derivative is larger than that displayed by $I_{\text{MATTER}}(z)$ are matter terms, with increasing density. The prior-free NN solution suggests therefore an “interacting dark sector”

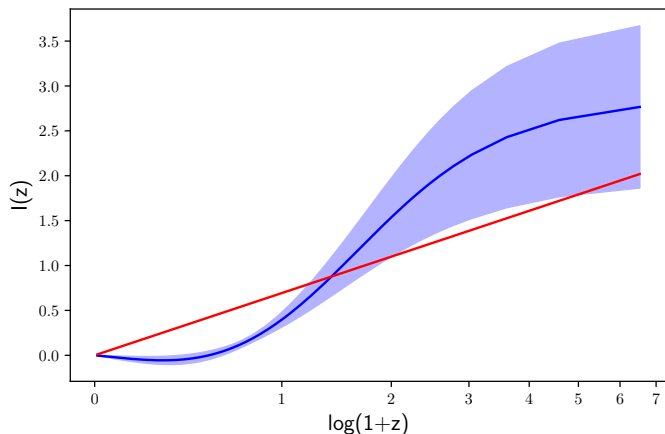


Fig. 3. Best-fit $I(z)$ from our NN regression (as in the upper panel of Fig. 2) in logarithmic scale, compared with the function $I_{\text{MATTER}}(z)$ obtained by assuming $w(z) \equiv 0$ in Eq. (3). The redshift intervals where the derivative of $I(z)$ is higher than that of $I_{\text{MATTER}}(z)$ represent “matter-like” contributions, while intervals with a lower derivative refer to energy-like contributions.

scenario, where a matter component decreases with time and, correspondingly, a dark energy component rises. This interpretation is also consistent with the nearly constant loss function for $\Omega_{\text{M}} < 0.3$: choosing values larger than 0.3 worsens the agreement because this amounts to overestimating the total matter component at $z \sim 0$. Values smaller than 0.3 can be compensated for by the matter component in $I(z)$. This interpretation is consistent with recent claims of an increasing value of Ω_{M} with redshift within a Λ CDM scenario (e.g., Colgáin et al. 2022).

4. Conclusions

Our conclusions are multifold. We have proposed and rigorously tested a NN approach for analyzing the Hubble diagram. The NN model-agnostic regression of the combined supernova and quasar catalog enables us to unequivocally reveal a strong tension with the “concordance” flat Λ CDM model. Finally, the analysis we carried out with the proposed NN approach

suggests an “interacting dark sector” scenario, where a dark matter component flows into dark energy, at least down to redshifts $z \sim 1.5$.

Acknowledgements. We acknowledge financial contribution from the agreement ASI-INAF n.2017-14-H.O. EL acknowledges the support of grant ID: 45780 Fondazione Cassa di Risparmio Firenze. FS acknowledges is financially supported by the National Operative Program (Programma Operativo Nazionale–PON) of the Italian Ministry of University and Research “Research and Innovation 2014–2020”, Project Proposals CIR01_00010. A sincere acknowledgement goes to Dr. Colasurdo, who first performed the data reduction of the LBT K_S spectra in her master thesis and developed the baseline analysis that we used as a benchmark. We also acknowledge Prof. Trakhtenbrot for kindly sharing the BASS data.

References

- Aviles, A., Bravetti, A., Capozziello, S., & Luongo, O. 2014, *Phys. Rev. D*, **90**, 043531
- Bargiacchi, G., Risaliti, G., Benetti, M., et al. 2021, *A&A*, **649**, A65
- Battye, R. A., Charnock, T., & Moss, A. 2015, *Phys. Rev. D*, **91**, 103508
- Capozziello, S., D’Agostino, R., & Luongo, O. 2020, *MNRAS*, **494**, 2576
- Cárdenas, V. H. 2015, *Phys. Lett. B*, **750**, 128
- Cattoën, C., & Visser, M. 2007, *Class. Quant. Grav.*, **24**, 5985
- Chevallier, M., & Polarski, D. 2001, *Int. J. Mod. Phys. D*, **10**, 213
- Colgáin, E., Sheikh-Jabbari, M. M., Solomon, R., Dainotti, M. G., & Stojkovic, D. 2022, arXiv e-prints [arXiv:2206.11447]
- Demianski, M., Piedipalumbo, E., Sawant, D., & Amati, L. 2017, *A&A*, **598**, A112
- Di Valentino, E., Melchiorri, A., & Silk, J. 2021, *ApJ*, **908**, L9
- Dialektopoulos, K., Said, J. L., Mifsud, J., Sultana, J., & Zarb Adami, K. 2022, *J. Cosmol. Astropart. Phys.*, **2022**, 023
- Heymans, C., Tröster, T., Asgari, M., et al. 2021, *A&A*, **646**, A140
- Holsclaw, T., Alam, U., Sansó, B., et al. 2010, *Phys. Rev. D*, **82**, 103502
- Liaw, R., Liang, E., Nishihara, R., et al. 2018, arXiv e-prints [arXiv:1807.05118]
- Lin, W., & Ishak, M. 2017, *Phys. Rev. D*, **96**, 023532
- Lusso, E., Piedipalumbo, E., Risaliti, G., et al. 2019, *A&A*, **628**, L4
- Lusso, E., Risaliti, G., Nardini, E., et al. 2020, *A&A*, **642**, A150
- Macaulay, E., Wehus, I. K., & Eriksen, H. K. 2013, *Phys. Rev. Lett.*, **111**, 161301
- Nunes, R. C., & Vagnozzi, S. 2021, *MNRAS*, **505**, 5427
- Colgáin, Ó., & E. & Sheikh-Jabbari, M. M., 2021, *Eur. Phys. J. C*, **81**, 892
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, *ApJ*, **876**, 85
- Risaliti, G., & Lusso, E. 2019, *Nat. Astron.*, **195**
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, **859**, 101
- Seikel, M., Clarkson, C., & Smith, M. 2012, *J. Cosmol. Astropart. Phys.*, **06**, 036
- Shafieloo, A., Kim, A. G., & Linder, E. V. 2012, *Phys. Rev. D*, **85**, 123530
- Wong, K. C., Suyu, S. H., Chen, G., et al. 2019, *MNRAS*, **498**, 1420

Appendix A: Data processing

Data come as a set $\mathcal{D} = \{(z^{(i)}, y^{(i)}, \Delta y^{(i)})\}$, with $i \in 1 \dots |\mathcal{D}|$. Each $y^{(i)}$ component is linked to d_L , the physical quantity of interest, as $y^{(i)} = 5 \log(d_L^{(i)}/10\text{pc})$. The first applied transformation is defined as follows:

$$y'^{(i)} = y^{(i)}/5 + 1, \quad \Delta y'^{(i)} = \Delta y^{(i)}. \quad (\text{A.1})$$

As such, data are traced back to the logarithm of the luminosity distance; every entry of the inspected data set is equal to $y^{(i)} = \log(d_L^{(i)})$.

Carrying out a first-order expansion of Eq. (1) in the main body of the paper, assuming a flat Universe ($\Omega_k \sim 0$) and inserting the expression of $H(z)$ as reported in the main text, yields

$$d_L = \alpha(z) \int_0^z dz' \left[\Omega_M (1+z')^3 + (1 - \Omega_M) e^{I(z')} \right]^{-\frac{1}{2}}, \quad (\text{A.2})$$

where $\alpha(z) = \frac{c(1+z)}{H_0}$. We proceeded by setting

$$y''^{(i)} = y'^{(i)} - \log(\alpha(z^{(i)})), \quad \Delta y''^{(i)} = \Delta y'^{(i)}. \quad (\text{A.3})$$

It is worth noting that the relative errors associated with c , z , and H_0 are negligible. Relation A.3 transforms into

$$y'''^{(i)} = 10y''^{(i)}, \quad \Delta y'''^{(i)} = 10y''^{(i)} \Delta y''^{(i)}. \quad (\text{A.4})$$

To simplify the notation, we drop the apex by setting $y''' \rightarrow y$ and obtain the sought connection between every $y^{(i)}$ and the function to be fitted $I(z)$, namely

$$y^{(i)} = \int_0^{z^{(i)}} dz' \left[\Omega_M (1+z')^3 + (1 - \Omega_M) e^{I(z')} \right]^{-\frac{1}{2}}. \quad (\text{A.5})$$

We point out that since $\arg \min_x f(x) = \arg \min_x C \cdot f(x)$, every manipulation that results in a constant factor in front of the loss function can be ignored. This is the case of every operation in the form of $\Delta y_{\text{new}}^{(i)} = C \cdot \Delta y_{\text{old}}^{(i)}$, which, indeed, results in a factor $\frac{1}{C^2}$ in front of the loss function (see Eq. (C.1)).

Appendix B: The employed neural network model

To approximate the nonlinear scalar function $I(z) : z \in \mathbb{R} \mapsto I(z) \in \mathbb{R}$, we made use of a so-called feed-forward architecture. The information flows from the input neuron, associated with $z^{(i)}$, to the output neuron, where the predicted value of $I_{\text{NN}}(z^{(i)})$ is displayed.

The transformation from layer k to its adjacent homolog, $k+1$, following a feed-forward arrangement, is characterized by two nested operations – (i) a linear map $W^{(k)} : \mathbb{R}^{N_k} \rightarrow \mathbb{R}^{N_{k+1}}$ and (ii) a nonlinear filter $\sigma^{(k+1)}(\cdot)$ – applied to each entry of the obtained vector. Here k ranges in the interval $1 \dots \ell$, where $N_1 = 1$ and ℓ is the number of layers (i.e., the depth of the NN). We have chosen $\sigma^{(k)} := \tanh$ for all $k < \ell - 1$, whereas $\sigma^{(\ell)} = \mathbb{I}$.

The activation of every neuron in layer k can be consequently obtained as

$$\mathbf{x}^{(k)} = W^{(k-1)}(\dots \sigma(W^{(2)}(\sigma(W^{(1)}z))) \dots).$$

Furthermore, we fixed $N_k = N_{k+1} \forall k \in 2 \dots \ell - 2$, meaning that every layer (except the first and the last) has the same size as the others. The size of the so-called hidden layer, N_2 , and the total

amount of layers, ℓ , are, consequently, the only hyperparameters to eventually be fixed.

Occasionally, a neuron-specific scalar, called bias, can be added after the application of each linear map, $W^{(k)}$. To allow for the solution $I_{\text{NN}}(0) = 0$ to be recovered, we set the bias to zero.

The output $I_{\text{NN}}(z)$ hence depends on $N = \sum_{k=1}^{\ell-1} N_k \times N_{k+1}$ free scalar parameters (the weights $W_{i,j}^{(k)}, i \in 1 \dots N_{k+1}, j \in 1 \dots N_k, k \in 1 \dots \ell - 1$), which constitute the target of the optimization.

Appendix C: Model optimization

The optimization described in this section was carried out by using parallel computing on GPUs (Liaw et al. 2018) and the minimization of the loss function as defined in the main text was performed via a variant of the SGD method, recalled below. First, data set \mathcal{D} is shuffled and divided into smaller subsets, \mathcal{B}_i , of size $|\mathcal{B}_i| = \beta$. These are the batches, and they meet the following condition: $\mathcal{D} = \sqcup_i^{N_b} \mathcal{B}_i$. Obviously, the number of batches, N_b , is equal to $\lceil \frac{|\mathcal{D}|}{\beta} \rceil$. The gradient with respect to every weight, W , entering the definition of the function L is computed, within each batch, as

$$G^{(i)} = \nabla_W L(W, \mathcal{B}_i) = \nabla_W \sum_{j: y^{(j)} \in \mathcal{B}_i} \left(\frac{y^{(j)} - y_{\text{pred}}^{(j)}(z^{(j)}; W)}{\Delta y^{(j)}} \right)^2. \quad (\text{C.1})$$

While i takes values in the range $1 \dots N_b$, the weights, W , are updated so as to minimize, via a stochastic procedure, the loss function. This is achieved as follows:

$$W \leftarrow W - \alpha_{lr} G^{(i)}. \quad (\text{C.2})$$

The hyperparameter α_{lr} is called the learning rate and drives the amount of stochasticity in the loss descent process. In the present work a more complex, yet conceptually equivalent, variant of the SGD called Adam is implemented.

A so-called epoch is completed when all batches have been used. The number of epochs, N_e , is another hyperparameter that has to be fixed a priori, as does the batch size, β . A high number of epochs (such as 400 or 600, as employed in the present application) is usually chosen. To avoid overfitting, the early stop technique is employed. Such a technical aid consists in taking a small subset, \mathcal{V} , of the data set ($\sim 15\%$ of \mathcal{D}) and excluding it from the training process. During training stages, hence, the employed data set is $\mathcal{D}' = \mathcal{D} - \mathcal{V}$. While applying SGD to the loss so as to minimize it, loss evaluation on data set \mathcal{V} , $L(I_{\text{NN}}, \mathcal{V})$ is also performed. When the latter function reaches a plateau, the optimization process is stopped. This procedure relies on two hyperparameters – δ , the absolute variation in L that can be considered a real loss change, and p , the number of consecutive epochs with no recorded variation before the fitting algorithm can eventually be terminated.

One additional hyperparameter needs to be mentioned: as already explained in the main body of the paper, the prediction y_{pred} involves a numerical integral of the NN approximating function, I_{NN} . The integration step dz' thus has to be set and underwent of a meticulous optimization. A hyper-optimization process designed to find the best set of hyperparameters was carried out, employing several CPL- and Λ CDM-like models. Such a process resulted in a set of parameters that were fixed and left unchanged during the trials. A list of the chosen hyperparameters is provided in Table C.1.

Table C.1. Hyperparameters employed.

| N_2 | ℓ | N_e | bs | α_{lr} | δ | p | dz' |
|-------|--------|-------|------|---------------|-----------|-----|-------------------|
| 20 | 5 | 600 | 100 | 10^{-6} | 10^{-6} | 35 | $5 \cdot 10^{-4}$ |

Appendix D: Simulation results

In the following we report the results of the regression model against the simulated settings mentioned, but not displayed, in the main text.

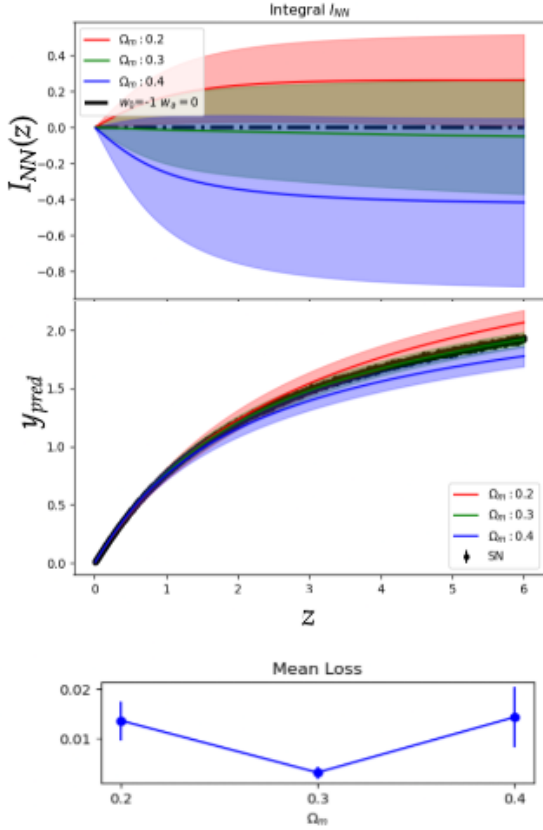


Fig. D.1. Simulations with a "perfect" sample (data set A): results of the NN analysis of a simulated sample of 4,000 objects with a log-flat redshift distribution and a negligible dispersion with respect to a flat Λ CDM model with $\Omega_M=0.3$. Top panel: Estimated values of $I(z)$ for different values of Ω_M (Eq. 3); the "correct" value for the simulated data is $I(z)\equiv 0$. Central panel: Hubble diagram with the reconstructed best-fit function obtained from the NN analysis. Bottom panel: Loss values for different values of Ω_M . The minimum is at $\Omega_M=0.3$, i.e., the "true" value. The corresponding $I(z)$ is consistent with zero at all redshifts. These results demonstrate that the NN analysis is able to recover the correct model and the "true" value of Ω_M .

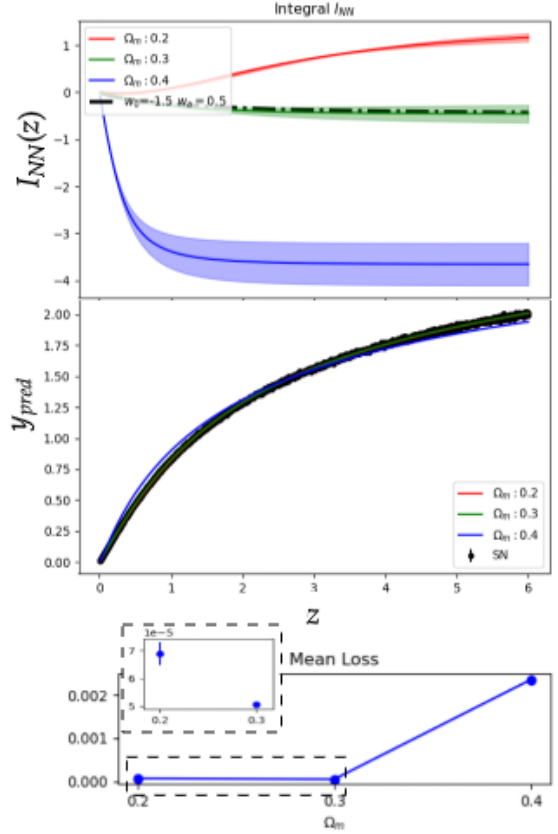


Fig. D.2. Results for data set B. The governing model is a CPL with $w_0 = -1.5, w_a = 0.5$.

Appendix E: Estimating the errors

To estimate the prediction error $\Delta y_{\text{pred}}(z)$, we employed a bootstrap method. To this end, the fitting procedure was arranged so as to produce B independent estimators of the quantity y_{pred} and $I_{\text{NN}}(z)$, namely $y_{\text{pred}}^{[k]}$ and $I_{\text{NN}}^{[k]}$ with $k \in 1 \dots B$. Each $y_{\text{pred}}^{[k]}$ is the result of an optimization process started from a subset $\mathcal{D}^{[k]} \subseteq \mathcal{D}$ obtained from \mathcal{D} via uniform sampling with a replacement of $|\mathcal{D}|$ elements. The prediction errors Δy_{pred} and ΔI_{NN} were then computed by extracting the standard deviation from both sets as

$$\Delta y_{\text{pred}}(z) = \sum_{k=1}^B \sqrt{\frac{(\bar{y}_{\text{pred}}(z) - y_{\text{pred}}^{[k]}(z))^2}{B-1}}, \quad (\text{E.1})$$

$$\Delta I_{\text{NN}}(z) = \sum_{k=1}^B \sqrt{\frac{(\bar{I}_{\text{NN}}(z) - I_{\text{NN}}^{[k]}(z))^2}{B-1}},$$

where symbols $\bar{y}_{\text{pred}}(z)$ and $\bar{I}_{\text{NN}}(z)$ represent the arithmetic mean of the estimates $y_{\text{pred}}^{[k]}$ and $I_{\text{NN}}^{[k]}$. Throughout this work, the errors were computed after $B = 80$ bootstrap samples.

Here we comment on the derivation of the indicator to gauge the correspondence of the fitted model with a conventional Λ CDM scheme. We began by formally expressing δy_{pred} , the distance of the obtained prediction with respect to the reference Λ CDM model, as

$$\delta y_{\text{pred}}^{\Lambda\text{CDM}}(I_{\text{NN}}; z) = \left. \frac{\delta y_{\text{pred}}}{\delta I} \right|_{I=\Lambda\text{CDM}} \delta I, \quad (\text{E.2})$$

$$= \left. \frac{\delta y_{\text{pred}}}{\delta I} \right|_{\Lambda\text{CDM}} (I_{\text{NN}} - I_{\Lambda\text{CDM}})$$

where $\frac{\delta y_{\text{pred}}}{\delta I}$ stands for the functional derivative and $I_{\Lambda\text{CDM}} = 0$. This equation can be further expanded to yield

$$\frac{\delta y_{\text{pred}}}{\delta I} \Big|_{I=\Lambda\text{CDM}} = -\frac{1}{2} \int_0^z \alpha(z')^{-\frac{3}{2}} (1 - \Omega_m) e^{I(z')} \Big|_{I=0}, \quad (\text{E.3})$$

where $\alpha(z') = \Omega_M (1 + z')^3 + (1 - \Omega_M) e^{I(z')}$. By eventually setting $\delta I = I_{\text{NN}}$, one gets

$$\delta y_{\text{pred}}^{\Lambda\text{CDM}}(I_{\text{NN}}; z) = \frac{\Omega_m - 1}{2} \int_0^z (\alpha(z')|_{I=0})^{-\frac{3}{2}} I_{\text{NN}}(z') dz'. \quad (\text{E.4})$$

We are finally in a position to introduce the scalar indicator that fulfills the purpose of quantifying the sought distance, normalized to the associated error. This is denoted as $\Delta_{\Lambda\text{CDM}}$ and takes the form

$$\Delta_{\Lambda\text{CDM}}(\mathcal{D}, I_{\text{NN}}) = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} \frac{\delta y_{\text{pred}}^{\Lambda\text{CDM}}(I_{\text{NN}}; z^{(i)})}{\Delta y_{\text{pred}}(I_{\text{NN}}; z^{(i)})}. \quad (\text{E.5})$$

The fitted integral function I_{NN} is deemed compatible with the ΛCDM model if $\Delta_{\Lambda\text{CDM}} < 1$. When this condition holds true, the predictions deviate from a ΛCDM by an amount that, on average, is smaller than the corresponding prediction error.

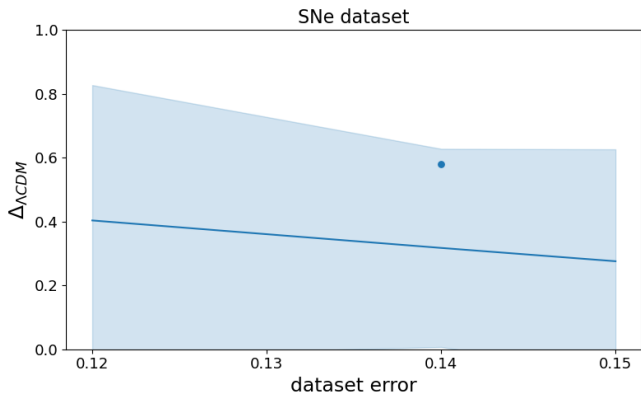


Fig. E.1. $\Delta_{\Lambda\text{CDM}}$ vs. the imposed error for the Pantheon data set (i.e., just supernovae). The dot stands for the experimental data, while the solid line and the shadowed region refer to the corresponding theoretical benchmarks, obtained as described in the text.

The indicator in Eq. (E.5) was computed for different mock samples, mimicking ΛCDM , with progressively increasing errors sizes, assumed uniform across data points, and values of Δy that range from zero to 0.15, thus including the value that is believed to apply to real data: ~ 0.14 . For every choice of the assigned error, 30 mock samples with $\Omega_m = 0.3$ were generated and subsequently fitted, assuming different choices of Ω_m , namely $\{0.2, 0.3, 0.4\}$. For every selected Ω_m , a bootstrap procedure was implemented (see the SI) to estimate y_{pred} , Δy_{pred} , and I_{NN} , ΔI_{NN} . The best-fit values are selected to be those associated with the smaller mean loss functions (evaluated against the imposed Ω_m). Following this choice, the mean and the variance of $\Delta_{\Lambda\text{CDM}}$ were computed from the outcomes of the fits and performed on the corresponding (30) independent realizations.

In Figs. E.1 to E.4, the results of the analysis for the different data sets are displayed. In Fig. E.1, solely supernova data ($z < 2$) were considered when carrying out the regression. The dot refers to the experimental data set (Lusso et al. 2020) and is set to the estimated error (0.14). It falls within the shadowed domain, thus



Fig. E.2. $\Delta_{\Lambda\text{CDM}}$ vs. the imposed error for the combined supernova + quasar sample at redshifts $z < 2$ (left panel) and $z > 2$ (right panel). The dots refer to the experimental data, while the solid lines and the shadowed regions stand for the corresponding theoretical benchmarks, obtained as described in the text.

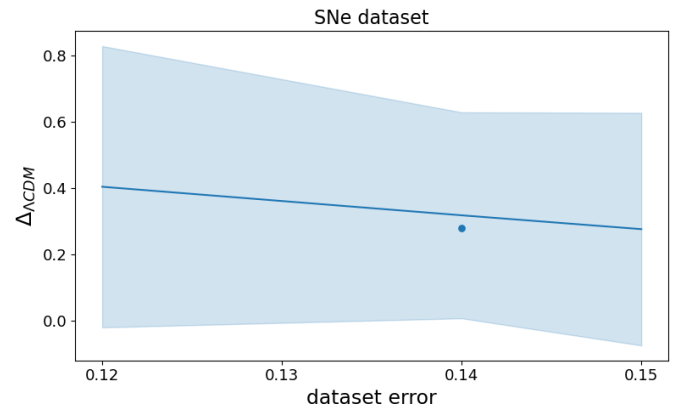


Fig. E.3. $\Delta_{\Lambda\text{CDM}}$ vs. the imposed error for the Pantheon data set (i.e., just supernovae). The reference mean and variance (represented as a shaded region) are shown in blue. The dot is obtained by processing the synthetic example generated via the CPL model.



Fig. E.4. $\Delta_{\Lambda\text{CDM}}$ vs. the imposed error for the combined supernova + quasar sample at redshifts $z < 2$ (left panel) and $z > 2$ (right panel). The reference mean and variance (represented as a shaded region) obtained with mock ΛCDM samples are shown in blue. The dots are obtained by processing the synthetic example generated via the CPL model.

implying that the examined data set is compatible with a ΛCDM model.

In Fig. E.2 we analyze the full data set (Pantheon + quasars). The regression was hence carried out by considering data spanning the whole range in z . After the fitting was performed, data were split into two different regions: at low ($z \leq 2$) or high ($z \geq 2$) redshifts. The dots refer to the experimental data set and are set to the estimated error (0.14). The portion of the data set

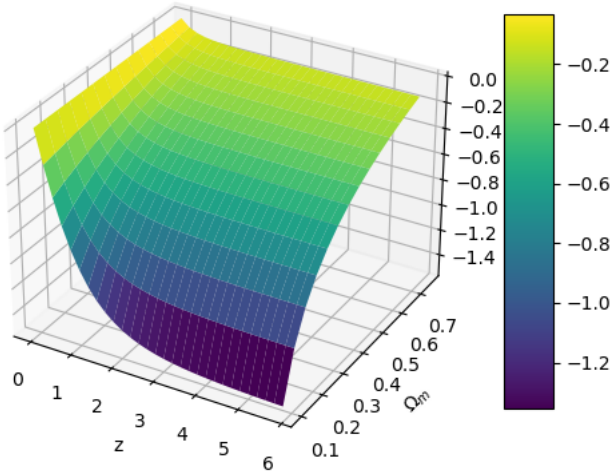
Functional derivative evaluated in Λ CDM

Fig. E.5. Plot of the functional derivative computed with Eq. (E.2) and varying Ω_m and z .

at low redshifts (mostly populated by supernovae) is compatible with a Λ CDM model with $\Omega_m = 0.3$, within statistical errors. The agreement is even more pronounced when the regression is carried out solely accounting for supernovae (see Fig. E.1).

Conversely, for $z > 2$, the point computed after available experiments, notably experiments involving quasars, is at a distance of about 5σ from the expected value of the indicator $\Delta_{\Lambda\text{CDM}}$. Hence, accounting for quasars enables us to conclude that the Λ CDM model is indeed extremely unlikely.

In Figs. E.3 and E.4 we repeat the analysis by employing a data set generated from a CPL model, with an error compatible with that estimated from experiments (equivalent to data sets D and F). The results indicate that accounting for data at high redshifts is mandatory for resolving the degeneracy between distinct generative models.

As a final point we elaborate on the reason why different models appear indistinguishable at low z . Function I_{NN} is the argument of a functional that goes from the space of function I to the space of the predictions. The way these two spaces communicate (or rather how function I reverberates on every y_{pred}) is a nontrivial function of the hyperparameters (e.g., Ω_m and the integration steps) and the domain explored. To clarify this point,

we plot the functional derivative $\frac{\delta y_{\text{pred}}}{\delta I}$ (evaluated at the Λ CDM model) against Ω_m and z . Via a visual inspection of Fig. E.5, the relevant impact of a low z and a large Ω_m is evident. The functional derivative is hence very small for the portion of the data set that is populated by the vast majority of supernova entries. This implies that different models (in terms of the associated $I(z)$) can yield very similar predictions. It is hence difficult to draw conclusions about the validity of different models if one solely deals with data at low redshifts.