# E-Value Analysis of DESI Data:
# A Complete Walkthrough

Mathematics, Data, Processing, Testing, and Assumptions

Jinyoung Kim

February 2026

**Abstract**

This document walks through every step of our e-value analysis of DESI (Dark Energy Spectroscopic Instrument) data, from the underlying cosmological physics to the final statistical conclusions. For each step, we explain *why* we do it, *what* the math is, and give a concrete *worked example* with real numbers. The goal is that a reader with undergraduate-level mathematics can follow the entire chain: cosmological distances → BAO measurements → statistical framework (e-values) → results and their meaning. We are explicit about every assumption made and every limitation encountered.

## Contents

# Part I
# The Physics

## 1 The Expanding Universe

**The Core Idea**

The universe is expanding. Galaxies are moving apart from each other, not because they are flying through space, but because space itself is stretching. The rate of this stretching depends on what the universe is made of.

### 1.1 Redshift

When a galaxy emits light at some time in the past, the light's wavelength gets stretched as the universe expands while the light travels to us. The **redshift** $z$ quantifies this stretching:

$$1 + z = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{1}{a} \tag{1}$$

where $a$ is the **scale factor** of the universe at the time the light was emitted ($a = 1$ today).

**Example: Redshift**

A galaxy at $z = 1$ emitted its light when the universe was half its current size ($a = 1/2$). Light from a $z = 0.5$ galaxy was emitted when the universe was $2/3$ its current size. Higher $z$ means further back in time, smaller universe.

### 1.2 The Hubble Parameter

The expansion rate is described by the **Hubble parameter**:

$$H(z) = H_0 \cdot E(z) \tag{2}$$

where $H_0 \approx 67.7$ km/s/Mpc is today's expansion rate and $E(z)$ is the **dimensionless Hubble parameter** that encodes how the expansion rate changes with redshift.

### 1.3 The Friedmann Equation

General relativity tells us $E(z)$ depends on the contents of the universe:

$$\boxed{E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_r(1+z)^4 + \Omega_k(1+z)^2 + \Omega_{DE}(z)}} \tag{3}$$

Each term represents a component of the universe:

| **Component** | **What it is** | **Symbol** | **Value (Planck 2018)** |
|---|---|---|---|
| Matter | Dark matter + baryons | $\Omega_m$ | 0.3111 |
| Radiation | Photons + neutrinos | $\Omega_r$ | $9 \times 10^{-5}$ |
| Curvature | Spatial geometry | $\Omega_k$ | $\approx 0$ |
| Dark energy | Accelerating expansion | $\Omega_{DE}$ | 0.6889 |

**Example: Computing $E(z)$ for $\Lambda$CDM**

At $z = 0.7$ with the standard $\Lambda$CDM model ($\Omega_{DE}$ is constant):

$$E(0.7) = \sqrt{0.3111 \times 1.7^3 + 9 \times 10^{-5} \times 1.7^4 + 0 + 0.6889} \tag{4}$$
$$= \sqrt{0.3111 \times 4.913 + 0.0007 + 0.6889} \tag{5}$$
$$= \sqrt{1.528 + 0.0007 + 0.6889} \tag{6}$$
$$= \sqrt{2.218} = 1.489 \tag{7}$$

So at $z = 0.7$, the expansion rate is about 1.49 times today's rate.

**Caution: Assumption: Flat Universe**

We assume $\Omega_k = 0$ (spatially flat universe) throughout. This is well-supported by CMB data but is nonetheless an assumption. If the universe has slight curvature, distance calculations would change.

# 2 Dark Energy: The Central Question

## 2.1 The Cosmological Constant ($\Lambda$CDM)

In the simplest model, dark energy is Einstein's **cosmological constant** $\Lambda$ with a fixed energy density. This is characterized by an **equation of state parameter** $w = -1$:

$$P_{DE} = w\rho_{DE}c^2 \quad \text{with } w = -1 \text{ (constant)} \tag{8}$$

Under $\Lambda$CDM, $\Omega_{DE}(z) = \Omega_{DE,0} = 0.6889$ — it does not change with redshift.

## 2.2 Dynamic Dark Energy ($w_0w_a$CDM)

What if dark energy is not constant? The CPL (Chevallier–Polarski–Linder) parametrization allows $w$ to evolve:

$$\boxed{w(a) = w_0 + w_a(1 - a) = w_0 + w_a\frac{z}{1+z}} \tag{9}$$

This gives two free parameters:

- $w_0$: the value of $w$ today ($z = 0$, $a = 1$)
- $w_a$: how much $w$ changes over time

The dark energy density then evolves as:

$$\Omega_{DE}(z) = \Omega_{DE,0} \cdot (1 + z)^{3(1+w_0+w_a)} \cdot \exp\left(-3w_a\frac{z}{1+z}\right) \tag{10}$$

**Example: DESI's Best-Fit Dynamic Dark Energy**

DESI DR2 finds $w_0 \approx -0.75$, $w_a \approx -1.05$ as the best fit. Let's see what $w$ looks like at different epochs:

| $z$ | $a = 1/(1+z)$ | $w(a) = -0.75 + (-1.05)(1-a)$ | Meaning |
|---|---|---|---|
| 0 | 1.0 | $-0.75$ | Today: slightly less negative than $-1$ |
| 0.5 | 0.667 | $-1.10$ | More negative than $\Lambda$ |
| 1.0 | 0.5 | $-1.28$ | Even more negative |
| 2.0 | 0.333 | $-1.45$ | Strongly phantom-like |

Under this model, dark energy was *stronger* in the past ($w < -1$, "phantom") and is weakening toward $w \to -0.75$ today. ΛCDM has $w = -1$ at all times.

**Caution: Assumption: CPL Parametrization**

We assume dark energy dynamics (if any) can be captured by two numbers $(w_0, w_a)$. This is a convenient but arbitrary choice. More complex evolution patterns would require different parametrizations.

# 3 Cosmological Distances

These are the quantities we actually compute and compare against data. All distances depend on $H(z)$, and therefore on the dark energy model.

## 3.1 Hubble Distance $D_H(z)$

The distance light would travel if the universe were expanding at the *instantaneous* rate at redshift $z$:

$$D_H(z) = \frac{c}{H(z)} = \frac{c}{H_0 E(z)} \tag{11}$$

This measures the expansion rate *at* redshift $z$ directly.

**Example: Hubble Distance**

At $z = 0.7$ under ΛCDM (we computed $E(0.7) = 1.489$):

$$D_H(0.7) = \frac{299792.458}{67.66 \times 1.489} = \frac{299792.458}{100.73} = 2976 \text{ Mpc} \tag{12}$$

## 3.2 Comoving Distance $D_C(z)$ and Transverse Comoving Distance $D_M(z)$

The total comoving distance to redshift $z$ is an integral over the whole line of sight:

$$D_C(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \tag{13}$$

For a flat universe, $D_M = D_C$. This measures the total accumulated distance, sensitive to the expansion history at all redshifts between 0 and $z$.

## 3.3 Volume-Averaged Distance $D_V(z)$

For measurements that average over angles (isotropic BAO):

$$D_V(z) = \left[z \cdot D_H(z) \cdot D_M(z)^2\right]^{1/3} \tag{14}$$

This combines the radial ($D_H$) and transverse ($D_M$) distances.

### 3.4 The Sound Horizon $r_d$

Before the universe was 380,000 years old, photons and baryons formed a hot plasma with sound waves propagating through it. When the universe cooled enough for atoms to form ("recombination"), these waves froze in place. The distance the waves traveled is the **sound horizon**:

$$r_d = \int_{z_{\rm rec}}^{\infty} \frac{c_s(z)}{H(z)}\, dz \approx 147 \text{ Mpc} \tag{15}$$

**Why $r_d$ Matters**

$r_d \approx 147$ Mpc is a *known physical length* calibrated by well-understood early-universe physics. It acts as a "standard ruler." By measuring how big this ruler *appears* at different redshifts, we can infer distances. This is what BAO measures.

### 3.5 What We Actually Compare to Data

DESI measures distance *ratios* scaled by the sound horizon:

$$\frac{D_M(z)}{r_d}, \qquad \frac{D_H(z)}{r_d}, \qquad \frac{D_V(z)}{r_d} \tag{16}$$

These ratios are what our code computes (in `cosmology.py`) and compares to the measured values.

**Example: Full Distance Calculation at $z = 0.706$**

Under $\Lambda$CDM ($w_0 = -1, w_a = 0$), with $r_d = 147.09$ Mpc:
$D_H(0.706)/r_d$: We need $E(0.706) \approx 1.497$, so:

$$\frac{D_H}{r_d} = \frac{c/(H_0 \cdot E(z))}{r_d} = \frac{299792.458/(67.66 \times 1.497)}{147.09} = \frac{2960}{147.09} \approx 20.12$$

$D_M(0.706)/r_d$: Requires numerical integration $\int_0^{0.706} dz'/E(z')$. The result is $D_M/r_d \approx 17.86$.
The DESI DR2 measurements at $z = 0.706$ are: $D_M/r_d = 17.35 \pm 0.18$, $D_H/r_d = 19.46 \pm 0.33$.

# Part II
# The Data

## 4 DESI DR2 BAO Measurements

**Data: Source and Format**

**Source:** Official DESI public data release via `CobayaSampler/bao_data` on GitHub.
**Files:**

- `desi_gaussian_bao_ALL_GCcomb_mean.txt`: 13 measurements
- `desi_gaussian_bao_ALL_GCcomb_cov.txt`: $13 \times 13$ covariance matrix

### 4.1 The Data Vector

The complete DESI DR2 dataset consists of 13 BAO measurements at 7 effective redshifts:

| Index | $z_{\text{eff}}$ | Quantity | Value | Tracer |
|---|---|---|---|---|
| 1 | 0.295 | $D_V/r_d$ | 7.942 | BGS |
| 2 | 0.510 | $D_M/r_d$ | 13.588 | LRG1 |
| 3 | 0.510 | $D_H/r_d$ | 21.863 | LRG1 |
| 4 | 0.706 | $D_M/r_d$ | 17.351 | LRG2 |
| 5 | 0.706 | $D_H/r_d$ | 19.455 | LRG2 |
| 6 | 0.934 | $D_M/r_d$ | 21.576 | LRG3+ELG1 |
| 7 | 0.934 | $D_H/r_d$ | 17.641 | LRG3+ELG1 |
| 8 | 1.321 | $D_M/r_d$ | 27.601 | ELG2 |
| 9 | 1.321 | $D_H/r_d$ | 14.176 | ELG2 |
| 10 | 1.484 | $D_M/r_d$ | 30.512 | QSO |
| 11 | 1.484 | $D_H/r_d$ | 12.817 | QSO |
| 12 | 2.330 | $D_H/r_d$ | 8.632 | Ly$\alpha$ |
| 13 | 2.330 | $D_M/r_d$ | 38.989 | Ly$\alpha$ |

### 4.2 Understanding the Tracers

DESI uses different types of objects to measure BAO at different redshifts:

| Tracer | Redshift Range | What it is |
|---|---|---|
| BGS | $z \sim 0.3$ | Bright Galaxy Survey: nearby bright galaxies |
| LRG | $z \sim 0.5$–$0.9$ | Luminous Red Galaxies: massive, red galaxies |
| ELG | $z \sim 0.9$–$1.3$ | Emission Line Galaxies: star-forming galaxies |
| QSO | $z \sim 1.5$ | Quasars: active galactic nuclei |
| Ly$\alpha$ | $z \sim 2.3$ | Lyman-$\alpha$ forest: absorption in quasar spectra |

### 4.3 The Covariance Matrix

The $13 \times 13$ covariance matrix $C$ encodes both measurement uncertainties and correlations. It is **block-diagonal**: measurements at different redshifts are uncorrelated, but $D_M/r_d$ and $D_H/r_d$ at the *same* redshift are correlated (typically anti-correlated).

**Example: Reading the Covariance Matrix**

At $z = 0.706$ (indices 4 and 5 in the data vector), the covariance block is:

$$C_{z=0.706} = \begin{pmatrix} 0.0324 & -0.0237 \\ -0.0237 & 0.1115 \end{pmatrix}$$

This tells us:

- $\sigma(D_M/r_d) = \sqrt{0.0324} = 0.180$ (1.0% precision)
- $\sigma(D_H/r_d) = \sqrt{0.1115} = 0.334$ (1.7% precision)
- Correlation: $\rho = \frac{-0.0237}{\sqrt{0.0324 \times 0.1115}} = -0.39$ (anti-correlated)

The anti-correlation is physical: if the BAO peak is measured to be at a slightly larger angle (larger $D_M$), the corresponding radial measurement ($D_H$) tends to be slightly smaller.

**Caution: Assumption: Gaussian Errors**

We assume the likelihood is multivariate Gaussian: $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) \propto \exp\big(-\frac{1}{2}\chi^2\big)$. This is standard for BAO summary statistics and validated by DESI, but it is an approximation. Non-Gaussianity could affect tail probabilities.

# Part III
# The Statistical Framework

## 5 Hypothesis Testing: The Setup

**What We Are Testing**

- $H_0$ (**Null**): The universe has a cosmological constant. $w_0 = -1$, $w_a = 0$.
- $H_1$ (**Alternative**): Dark energy is dynamic. $w_0$ and $w_a$ are free parameters.

We want to know: does the DESI data provide compelling evidence that $H_0$ is wrong and $H_1$ is better?

### 5.1 The $\chi^2$ Statistic

The fundamental measure of fit quality is the chi-squared statistic:

$$\boxed{\chi^2 = (\mathbf{d} - \mathbf{t})^T \, C^{-1} \, (\mathbf{d} - \mathbf{t})} \tag{17}$$

where:

- $\mathbf{d}$ is the data vector (13 measurements)
- $\mathbf{t}(\boldsymbol{\theta})$ is the theory prediction vector (depends on cosmological parameters)
- $C$ is the $13 \times 13$ covariance matrix
- $C^{-1}$ is its inverse

Smaller $\chi^2$ means better fit. The improvement of $H_1$ over $H_0$ is:

$$\Delta\chi^2 = \chi^2_{H_0} - \chi^2_{H_1} \tag{18}$$

**Example: $\chi^2$ With Real Numbers**

For DESI DR2, the results are:

$$\chi^2_{\Lambda\mathrm{CDM}} \approx 25.4 \quad \text{(13 data points, 0 free params)} \tag{19}$$
$$\chi^2_{w_0 w_a\mathrm{CDM}} \approx 13.5 \quad \text{(13 data points, 2 free params)} \tag{20}$$
$$\Delta\chi^2 = 25.4 - 13.5 = 11.9 \tag{21}$$

With 2 extra parameters and $\Delta\chi^2 = 11.9$, a naive frequentist interpretation gives $\sqrt{\Delta\chi^2} \approx 3.5\sigma$.
**But is this real evidence, or just overfitting?** This is where e-values come in.

### 5.2 The Log-Likelihood

Under Gaussian errors, the log-likelihood is:

$$\ln \mathcal{L} = -\frac{1}{2}\chi^2 - \frac{1}{2}\ln|C| - \frac{n}{2}\ln(2\pi) \tag{22}$$

The last two terms are constants (they don't depend on the model), so differences in log-likelihood are determined entirely by $\Delta\chi^2$:

$$\ln \mathcal{L}_{H_1} - \ln \mathcal{L}_{H_0} = -\frac{1}{2}(\chi^2_{H_1} - \chi^2_{H_0}) = \frac{\Delta\chi^2}{2} \tag{23}$$

## 6 E-Values: What They Are and Why We Use Them

### 6.1 The Problem with Just Using $\Delta\chi^2$

Why not just report $\Delta\chi^2 = 11.9$ and declare victory?

**Caution: The Overfitting Problem**

When we fit $(w_0, w_a)$ to the same data we use to evaluate $\Delta\chi^2$, the improvement is *guaranteed* to be positive even if $H_0$ is true. Two free parameters can always improve the fit by chance. This inflates $\Delta\chi^2$ and makes the evidence look stronger than it really is.
A correction factor exists (Wilks' theorem says $\Delta\chi^2$ should follow a $\chi^2$ distribution with 2 degrees of freedom under $H_0$), but this relies on regularity conditions that may not hold here, and it doesn't address whether the *specific* parameter values generalize.

### 6.2 Definition of an E-Value

**Definition 1** (E-Value)**.** *A non-negative random variable $E$ is an **e-value** for testing $H_0$ if:*

$$\boxed{\mathbb{E}[E \mid H_0] \leq 1} \tag{24}$$

*That is, the expected value of $E$ under the null hypothesis is at most 1.*

**Intuition**

Think of $E$ as "evidence multiplied." Under $H_0$:

- On average, $E \leq 1$ (you don't expect to accumulate false evidence)
- A large $E$ is surprising — it's unlikely under $H_0$
- By Markov's inequality: $\mathbb{P}_{H_0}(E \geq 1/\alpha) \leq \alpha$

So if $E = 100$, the probability of seeing $E \geq 100$ under $H_0$ is at most $1/100 = 1\%$.

### 6.3 E-Values vs. P-Values

| | **P-value** | **E-value** |
|---|---|---|
| **Definition** | $P$(more extreme data $\mid H_0$) | Random variable with $\mathbb{E}[E\|H_0] \leq 1$ |
| **Interpretation** | How surprising is the data? | How much evidence against $H_0$? |
| **Combining** | Complex (Fisher's method, etc.) | Simple: multiply (if independent) |
| **Optional stopping** | Invalidates the test | Still valid |
| **Overfitting risk** | High if model is fitted post-hoc | Can be controlled (data-splitting) |

### 6.4 Interpreting E-Values

| **E-value** | **Interpretation** |
|---|---|
| $E < 1$ | Evidence *favors* $H_0$ (null) |
| $E = 1$ | No evidence either way |
| $E \sim 3$ | Weak evidence against $H_0$ |
| $E \sim 10$ | Moderate evidence against $H_0$ |
| $E \sim 100$ | Strong evidence against $H_0$ |
| $E \sim 1000$ | Very strong evidence |

**Approximate sigma conversion:** For rough comparison with frequentist significance, one can use $\sigma \approx \sqrt{2 \ln E}$. For example, $E = 100$ gives $\sigma \approx 3.0$. But this conversion is approximate; e-values and p-values answer different questions.

## 7 Three Methods to Compute E-Values

We implement three increasingly careful methods. Understanding why each exists is essential.

### 7.1 Method 1: Simple Likelihood Ratio

**The Idea**

The most basic e-value: how much better does $H_1$ fit the data compared to $H_0$?

$$E = \frac{\mathcal{L}(\mathbf{d} \mid H_1)}{\mathcal{L}(\mathbf{d} \mid H_0)} = \exp\left(\frac{\Delta\chi^2}{2}\right) \tag{25}$$

**Why this is a valid e-value** (for *fixed* $H_1$):
Under $H_0$, data follows distribution $P_0$. Then:

$$\mathbb{E}_{H_0}[E] = \int \frac{P_1(x)}{P_0(x)} P_0(x)\, dx = \int P_1(x)\, dx = 1 \tag{26}$$

**Why it's dangerous in practice:**
If we first fit $(w_0, w_a)$ to the data and *then* compute this ratio, $H_1$ was chosen to maximize the likelihood on this specific data. The proof above requires $H_1$ to be fixed *before* seeing the data.

**Example: Simple LR E-Value (BIASED)**

With DESI DR2 and the DESI best-fit $w_0 = -0.75$, $w_a = -1.05$:

$$\Delta\chi^2 \approx 11.9 \tag{27}$$

$$E_{\text{simple}} = e^{11.9/2} = e^{5.95} \approx 384 \tag{28}$$

Looks very strong! But the alternative was fitted to the same data used for evaluation, so this is **biased and invalid**.

### 7.2 Method 2: GROW Mixture E-Value

**The Idea**

Instead of using one specific $(w_0, w_a)$, average the likelihood ratio over a grid of possible alternatives. This "pre-specifies" the alternative as a mixture, preventing cherry-picking.

$$E_{\text{mix}} = \int \frac{\mathcal{L}(\mathbf{d} \mid w_0, w_a)}{\mathcal{L}(\mathbf{d} \mid H_0)} \pi(w_0, w_a)\, dw_0\, dw_a \tag{29}$$

In practice, we discretize: define a $10 \times 10$ grid over $(w_0, w_a) \in [-1.5, -0.5] \times [-2.0, 1.0]$, compute the likelihood ratio at each grid point, and average (with uniform weights):

$$E_{\text{mix}} = \frac{1}{100} \sum_{i=1}^{100} \exp\left( \frac{\chi^2_{H_0} - \chi^2_i}{2} \right) \tag{30}$$

**Why this is valid:** Each grid point gives a valid e-value (since $H_1$ is specified before seeing data). The average of e-values is an e-value (by linearity of expectation).

**The problem: prior sensitivity.** The result depends on which grid range we use:

| **Prior Range** | $w_0$ range | $w_a$ range | **E-value** |
|---|---|---|---|
| Narrow | $[-1.2, -0.8]$ | $[-1.0, 0.5]$ | $\sim 97$ |
| Default | $[-1.5, -0.5]$ | $[-2.0, 1.0]$ | $\sim 15$ |
| Wide | $[-2.0, 0.0]$ | $[-3.0, 2.0]$ | $\sim 17$ |

The factor of $\sim$7 variation shows this method is not robust to prior choice.

### 7.3 Method 3: Data-Split E-Value

**The Idea — Our Primary Method**

Split the data into two parts. Use one part to train (fit the alternative), and the other part to test. Because the test data was never used for fitting, the resulting e-value is honest.

**Procedure:**

1. **Split**: Divide 13 measurements into training set (7 measurements at $z < 1$) and test set (6 measurements at $z \geq 1$).
2. **Train**: Fit $(w_0, w_a)$ by minimizing $\chi^2$ on the training data only.
3. **Test**: Compute the likelihood ratio on the test data using the fitted parameters.

$$\boxed{E_{\text{split}} = \frac{\mathcal{L}(D_{\text{test}} \mid \hat{w}_0, \hat{w}_a)}{\mathcal{L}(D_{\text{test}} \mid H_0)} = \exp\left(\frac{\chi^2_{\text{test},H_0} - \chi^2_{\text{test},H_1}}{2}\right)} \tag{31}$$

**Why this is valid:**

Conditional on $D_{\text{train}}$, the fitted parameters $(\hat{w}_0, \hat{w}_a)$ are fixed constants. From $D_{\text{test}}$'s perspective, $H_1$ is fully specified before the test data is observed. Therefore:

$$\mathbb{E}[E_{\text{split}} \mid D_{\text{train}}] = 1 \quad \Rightarrow \quad \mathbb{E}[E_{\text{split}}] = 1 \tag{32}$$

**Example: Data-Split E-Value (VALID)**

**Step 1: Split the data.**
Training set ($z < 1$): 7 measurements at $z = 0.295, 0.51, 0.51, 0.706, 0.706, 0.934, 0.934$.
Test set ($z \geq 1$): 6 measurements at $z = 1.321, 1.321, 1.484, 1.484, 2.33, 2.33$.
**Step 2: Extract sub-covariance matrices.**
$C_{\text{train}}$ is the $7 \times 7$ submatrix (rows/cols 1–7 of the full matrix).
$C_{\text{test}}$ is the $6 \times 6$ submatrix (rows/cols 8–13). Because the full covariance is block-diagonal across redshift bins, these blocks are independent.
**Step 3: Fit on training data.**
Minimize $\chi^2_{\text{train}}(w_0, w_a)$ using L-BFGS-B optimizer with bounds $w_0 \in [-2, 0]$, $w_a \in [-3, 2]$. Starting point: $w_0 = -0.9$, $w_a = -0.5$.
Result: $\hat{w}_0 = -0.78$, $\hat{w}_a = -0.52$.
**Step 4: Evaluate on test data.**
Compute $\chi^2_{\text{test}}$ under both models:

$$\chi^2_{\text{test, }\Lambda\text{CDM}} \approx 5.8 \tag{33}$$

$$\chi^2_{\text{test, }w_0 w_a} \approx 5.1 \tag{34}$$

$$\Delta\chi^2_{\text{test}} = 0.7 \tag{35}$$

**Step 5: Compute e-value.**

$$E_{\text{split}} = e^{0.7/2} = e^{0.35} \approx 1.4 \tag{36}$$

**Interpretation:** The alternative model predicts high-redshift data only 1.4 times better than ΛCDM. This is essentially no evidence — the signal does not generalize.

**Caution: Assumption: Independence of Training and Test Sets**

The validity of data-split e-values requires $D_{\text{train}}$ and $D_{\text{test}}$ to be independent. This holds here because the DESI covariance matrix is block-diagonal across redshift bins: measurements at different redshifts are uncorrelated. If there were cross-redshift correlations (e.g., from systematic effects), this assumption would be violated.

# Part IV
# The Processing Pipeline

## 8 Step-by-Step: What the Code Does

Here is the complete processing chain, corresponding to the code modules:

### 8.1 Step 1: Load Data (`data_loader.py`)

1. Read `desi_gaussian_bao_ALL_GCcomb_mean.txt`: parse each line as $(z, \text{value}, \text{quantity})$.
2. Read `desi_gaussian_bao_ALL_GCcomb_cov.txt`: load $13 \times 13$ matrix.
3. Infer tracer type from redshift (e.g., $z = 0.295 \to$ BGS).
4. Package into a `BAODataset` object with arrays for $z$, data, covariance, quantities, tracers.

**Assumption:** The files are unmodified from the official release.

### 8.2 Step 2: Compute Theory Predictions (`cosmology.py`)

For given cosmological parameters $(w_0, w_a)$:

1. Compute $E(z)$ using Eq. (3) with dark energy evolution from Eq. (10).
2. Compute $D_H(z) = c/[H_0 E(z)]$.
3. Compute $D_C(z) = \frac{c}{H_0} \int_0^z dz'/E(z')$ via numerical integration (`scipy.integrate.quad`).
4. Compute $D_M(z) = D_C(z)$ (flat universe).
5. Compute $D_V(z) = [z \cdot D_H \cdot D_M^2]^{1/3}$.
6. Divide by $r_d = 147.09$ Mpc.
7. Build theory vector matching the order of the data vector (which entries are $D_M/r_d$, which are $D_H/r_d$, which are $D_V/r_d$).

### 8.3 Step 3: Compute $\chi^2$ (`cosmology.py`)

$$\chi^2 = (\mathbf{d} - \mathbf{t})^T C^{-1} (\mathbf{d} - \mathbf{t}) \tag{37}$$

Implemented as: compute residual vector $\mathbf{r} = \mathbf{d} - \mathbf{t}$, invert covariance via `numpy.linalg.inv`, compute quadratic form.

**Caution: Assumption: Invertible Covariance**

We compute $C^{-1}$ directly. This works because $C$ is $13 \times 13$ and well-conditioned (condition number $\sim 10^3$). For larger matrices, one would use Cholesky decomposition for numerical stability.

### 8.4 Step 4: Fit the Alternative Model (`evalue_analysis.py`)

For the data-split method:

1. Extract training data subset and sub-covariance matrix.
2. Define objective: $\chi^2_{\text{train}}(w_0, w_a)$.
3. Minimize using L-BFGS-B with bounds $w_0 \in [-2, 0]$, $w_a \in [-3, 2]$.
4. Output: best-fit $\hat{w}_0, \hat{w}_a$.

### 8.5 Step 5: Compute E-Value (`evalue_analysis.py`)

For each method:

1. Compute theory vectors under $H_0$ and $H_1$.
2. Compute $\chi^2$ under both models.
3. Compute $E = \exp(\Delta\chi^2/2)$.

# Part V
# The Testing and Results

## 9 Main Results

### 9.1 Same-Data Analysis (For Reference Only)

| Method | E-value | $\sim\sigma$ | Valid? |
|---|---|---|---|
| Simple LR (same data) | 392 | 3.5 | **NO** — overfitted |
| GROW mixture (narrow) | 97 | 3.0 | Prior-sensitive |
| GROW mixture (default) | 15 | 2.3 | Prior-sensitive |
| GROW mixture (wide) | 17 | 2.4 | Prior-sensitive |

### 9.2 Data-Split Analysis (Valid)

| Split Strategy | $E_{\mathbf{split}}$ | $\sim\sigma$ |
|---|---|---|
| $z < 1$ train, $z \geq 1$ test | **1.4** | 0.8 |
| Alternating bins | 2.1 | 1.2 |
| Random 50/50 (10 trials avg.) | $1.2 \pm 0.8$ | $0.6 \pm 0.5$ |

**The Key Finding**

The **280× reduction** from $E = 392$ (biased) to $E = 1.4$ (valid) shows that the apparent evidence for dynamic dark energy does *not* generalize out of sample. Parameters fitted on low-$z$ bins do not predict high-$z$ bins better than ΛCDM.
Every split strategy gives $E < 3$.

### 9.3 Cross-Dataset Validation

We also test whether parameters from other experiments predict DESI data better than ΛCDM:

| Train on | Test on | $(w_0, w_a)$ | E-value | Interpretation |
|---|---|---|---|---|
| DESI (fitted) | Pantheon+ | $(-0.86, -0.43)$ | 1.5 | No evidence |
| DESI (fitted) | DES-Y5 | $(-0.86, -0.43)$ | 86 | Moderate |
| Pantheon+ | DESI | $(-0.90, -0.20)$ | 2049 | Strong |
| DES-Y5 | DESI | $(-0.65, -1.20)$ | **0.19** | *Favors ΛCDM!* |

**The DES-Y5 Puzzle**

When DES-Y5's best-fit parameters are used to predict DESI data, the prediction is *worse* than ΛCDM ($E = 0.19 < 1$). This means DES-Y5 and DESI are pulling in different directions — their preferred regions of $(w_0, w_a)$ space are incompatible. This "tension" between datasets may be driving the apparent signal when they are combined.

## 10 DR1 → DR2 Temporal Validation

DESI released DR1 (Year 1, 2024) and DR2 (Years 1–3, 2025). We test: do parameters fitted on DR1 predict DR2?

| **Scenario** | $\Delta\chi^2$ | **E-value** | **Note** |
|---|---|---|---|
| DR2 fit on DR2 | ∼12 | ∼400 | BIASED |
| DR1 fit predicting DR2 | varies | varies | Semi-valid* |

*__Caveat__: DR2 *contains* DR1 (DR2 is 3 years of data including Year 1). They are not independent, so this is a test of *stability* rather than true out-of-sample prediction.

# Part VI
# Assumptions, Limitations, and Caveats

## 11 Complete List of Assumptions

We organize every assumption into categories with an assessment of how critical each one is.

### 11.1 Cosmological Assumptions

1. **Flat universe** ($\Omega_k = 0$). *Impact: Low.* Well-supported by CMB.
2. **CPL parametrization** for dark energy ($w = w_0 + w_a(1 - a)$). *Impact: Medium.* A different parametrization could change results.
3. **Planck 2018 fiducial parameters** ($\Omega_m = 0.3111$, $h = 0.6766$, $r_d = 147.09$ Mpc). *Impact: Low.* Small changes in these don't qualitatively change results.
4. **Standard distance formulas** without full Boltzmann code (CAMB/CLASS). *Impact: Low–Medium.* Our distances agree with DESI's to $< 0.5\%$.

### 11.2 Statistical Assumptions

5. **Gaussian likelihood**. *Impact: Low.* Standard for BAO summary statistics; validated by DESI.
6. **Published covariance matrix is correct**. *Impact: Medium.* We rely entirely on DESI's error estimates.
7. **Independence of redshift bins** (block-diagonal covariance). *Impact: Medium.* This is critical for the data-split e-value validity. Cross-redshift systematics would violate this.
8. **Data-split e-value validity** (training and test sets are independent). *Impact: High.* The core assumption of our main result. Follows from assumption 7.

### 11.3 Methodological Assumptions

9. **BAO-only analysis** (no CMB, no supernovae). *Impact: High.* DESI's full 3–4$\sigma$ claim uses combined data. Our BAO-only analysis tests a weaker claim.

10. **Point estimates for supernova constraints** (in cross-dataset analysis). *Impact: Medium.* We use published best-fit $(w_0, w_a)$ rather than full posteriors.

11. **Optimizer convergence**. *Impact: Low.* L-BFGS-B with reasonable bounds reliably finds the minimum for this smooth, low-dimensional problem.

## 12 Known Limitations

1. **Reduced statistical power from data-splitting.** By using only 6 of 13 measurements for testing, we lose power. However, if the true signal were as strong as $\Delta\chi^2 \sim 12$, we would still expect $E \gg 1$ on the test set.

2. **Choice of split point.** We split at $z = 1$, but other splits give similar results ($E < 3$ for all).

3. **No systematic error budget.** We treat the covariance matrix as exact. Unknown systematics could broaden error bars or introduce biases.

4. **Simplified cross-dataset analysis.** Using point estimates for supernova constraints (rather than full likelihoods with their own covariance) is approximate.

5. **No model-averaging or Bayesian comparison.** We compare two specific models. There may be other parametrizations that better capture any real dark energy evolution.

## 13 Comparison to Other Analyses

| Analysis | Method | Finding | Conclusion |
|---|---|---|---|
| DESI DR2 (official) | Frequentist $\Delta\chi^2$ | 3–4$\sigma$ | Dynamic DE |
| Ong et al. (2025) | Bayesian evidence | $\ln B = -0.57$ | Favors $\Lambda$CDM |
| Wang & Mota (2025) | Tension metrics | $2.95\sigma$ tension | Datasets inconsistent |
| **This work** | **E-values** | $E = 1.4$ | **No robust evidence** |

Three independent approaches (Bayesian, tension metrics, e-values) all question the robustness of the frequentist 3–4$\sigma$ claim.

# Part VII
# Summary

## 14 What We Did

1. Loaded official DESI DR2 BAO data: 13 measurements of cosmic distances at 7 redshifts.

2. Computed theoretical distance predictions under $\Lambda$CDM ($w = -1$, no dark energy evolution) and $w_0w_a$CDM (dark energy evolves).

3. Used three e-value methods to assess evidence:
   - Simple likelihood ratio (biased, for reference)
   - GROW mixture (prior-sensitive)
   - Data-split (honest, our main result)
4. Performed cross-dataset validation using supernova constraints.
5. Checked DR1→DR2 temporal stability.

# 15 What We Found

1. The naive $\chi^2$ improvement ($\Delta\chi^2 = 11.9$) gives an apparent $\sim 3.5\sigma$ signal. But this is computed on the same data used for fitting.
2. The **data-split e-value is** $E = 1.4$: parameters fitted on low-$z$ bins do not predict high-$z$ bins better than $\Lambda$CDM.
3. The **280× drop** from $E = 392$ (biased) to $E = 1.4$ (valid) directly quantifies how much of the evidence is due to overfitting.
4. **Cross-dataset validation** reveals that DES-Y5's best-fit parameters predict DESI data *worse* than $\Lambda$CDM ($E = 0.19$), suggesting a tension between datasets rather than consistent new physics.
5. These findings align with independent Bayesian and tension analyses.

# 16 What This Means

The evidence for dynamical dark energy from DESI, while genuinely interesting, does not survive out-of-sample validation. This does *not* prove that dark energy is constant — it means we need more data. Key next steps include:

- DESI DR3+ ($\sim$2027): more data, smaller errors
- Resolution of inter-dataset tensions (especially DESI vs. DES-Y5)
- Independent confirmation from Euclid and the Roman Space Telescope

---

*Code and data:* `https://github.com/jinyoungkim927/desi-evalue-analysis`
*DESI data:* `https://github.com/CobayaSampler/bao_data`

J. Kim, February 2026