Exhaustive Symbolic Regression

Learning Astrophysics directly from Data

Harry Desmond

ICG Portsmouth

w/ Deaglan Bartlett & Pedro Ferreira

arXiv:2211.11461 arXiv:2301.04368 arXiv:2304.06333 arXiv:2310.16786

21 May 2024

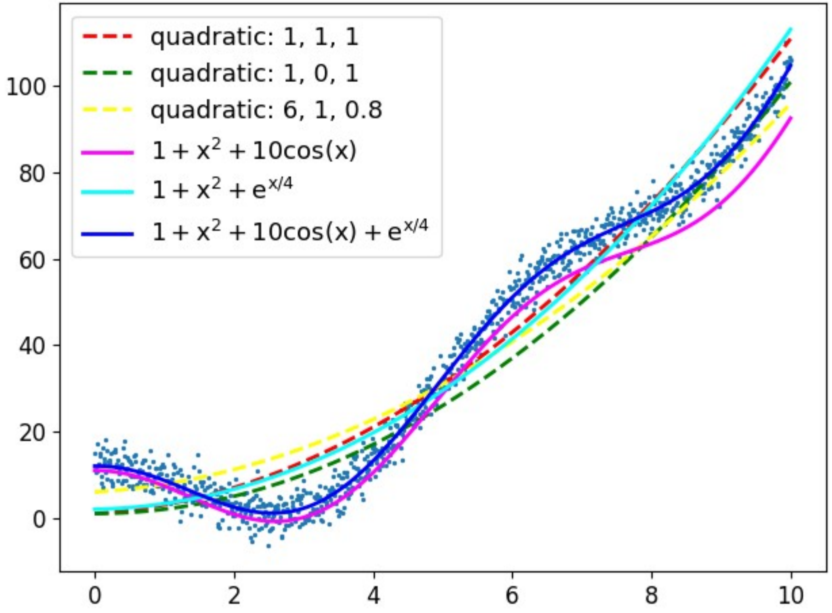
Cosmo21

Symbolic Regression overview

- Discover *functions* describing a dataset rather than parameters of predefined function

Numerical regression: $y = 6 + 1x + 0.8x^2$

Symbolic regression: $y = 1 + x^2 + 10\cos(x) + e^{x/4}$

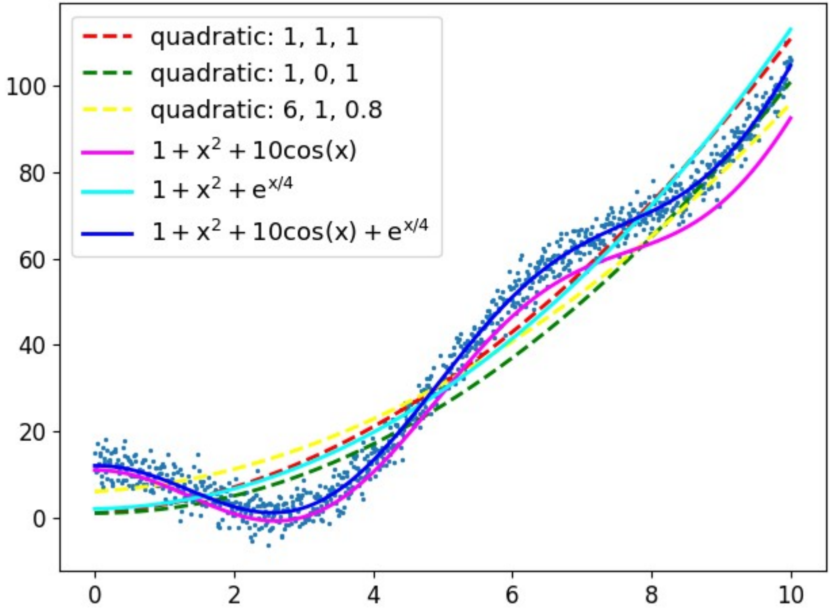


Symbolic Regression overview

- Discover *functions* describing a dataset rather than parameters of predefined function
- Advantages:
  - Much more general (reduces confirmation bias)
  - Highly interpretable
  - Easy to prevent overfitting
- Difficulties:
  - Larger search space makes convergence harder
  - Optimisation methods of numerical regression not applicable

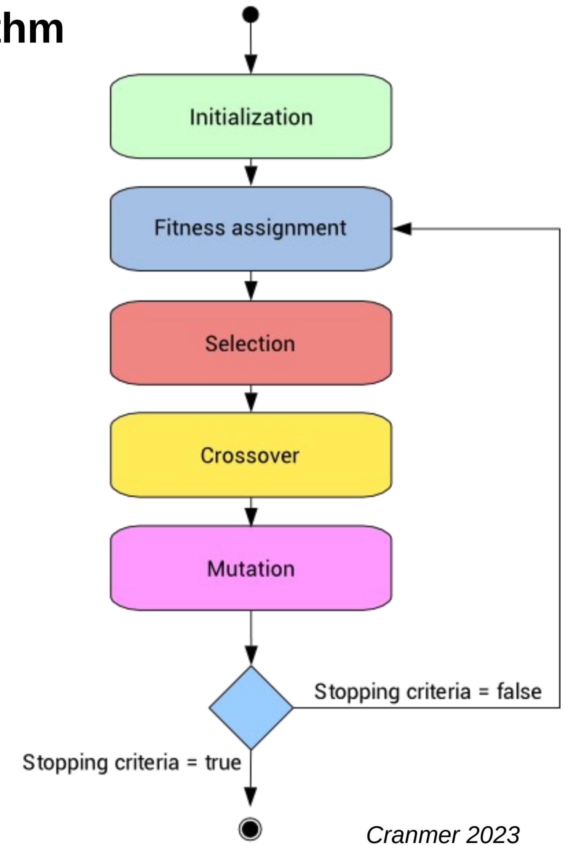
Numerical regression: $y = 6 + 1x + 0.8x^2$

Symbolic regression: $y = 1 + x^2 + 10\cos(x) + e^{x/4}$

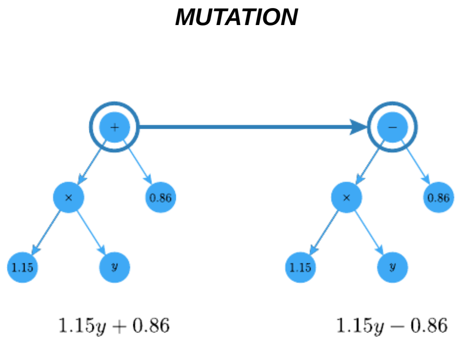


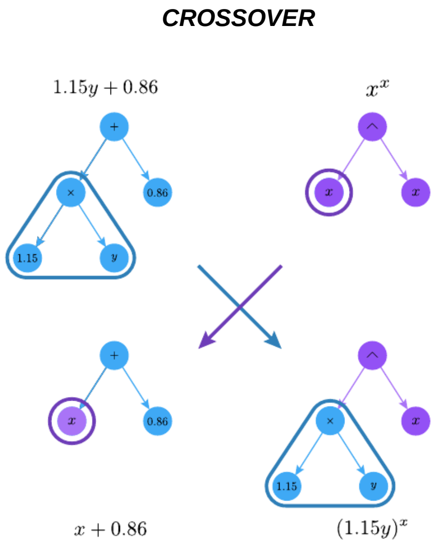
Traditional Symbolic Regression
I. Generating functions

Genetic Algorithm
(e.g. *PySR*, *DataModeler*)



Cranmer 2023





Traditional Symbolic Regression
II. Assessing functions

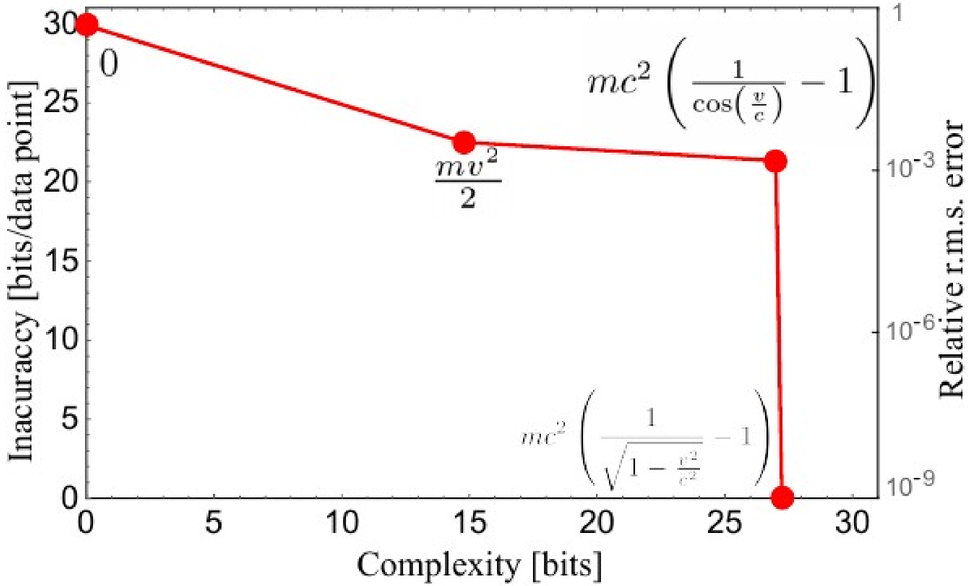
- Problem: Can typically get 0 error with some (very complex) overfitted function
- Solution: *two* objectives, accuracy and simplicity
- The best equations are the ones that cannot be made more accurate without also being made more complex

Traditional Symbolic Regression
II. Assessing functions

- Problem: Can typically get 0 error with some (very complex) overfitted function

- Solution: *two* objectives, accuracy and simplicity

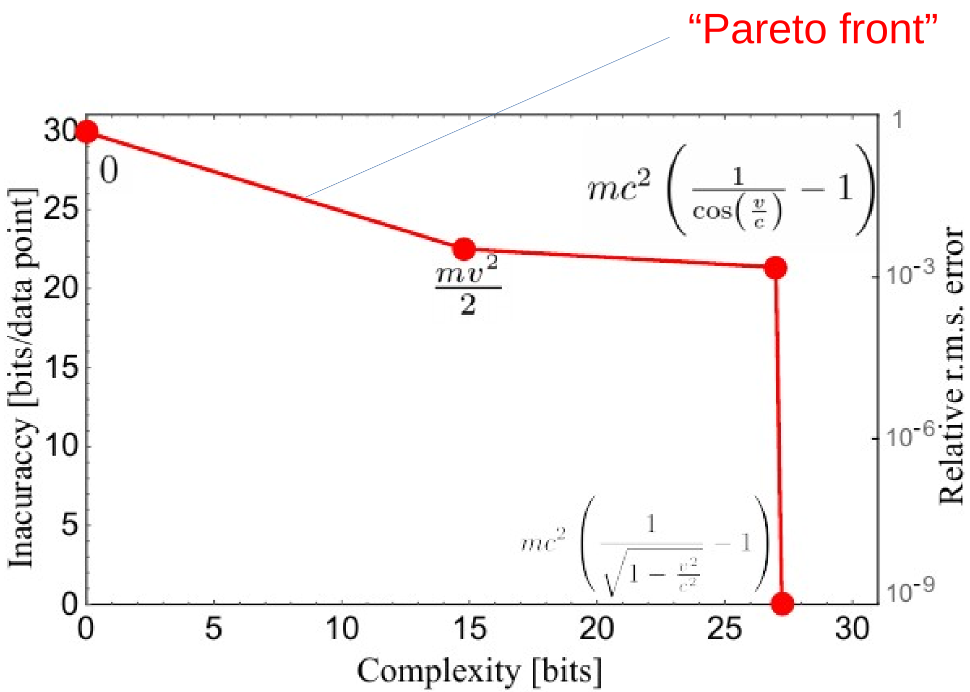
- The best equations are the ones that cannot be made more accurate without also being made more complex



Udrescu+ 2020

Traditional Symbolic Regression
II. Assessing functions

- Problem: Can typically get 0 error with some (very complex) overfitted function
- Solution: *two* objectives, accuracy and simplicity
- The best equations are the ones that cannot be made more accurate without also being made more complex ("Pareto-optimal")



Udrescu+ 2020

Exhaustive Symbolic Regression

Designed to overcome two problems:

1 Stochastic method may fail to find any given function

2 Typical accuracy definitions fail to account for data uncertainties, and complexity definition is largely arbitrary. The two are incommensurable.

Exhaustive Symbolic Regression

Designed to overcome two problems:

1 Stochastic method may fail to find any given function

$\Rightarrow$ Search exhaustively, complexity by complexity

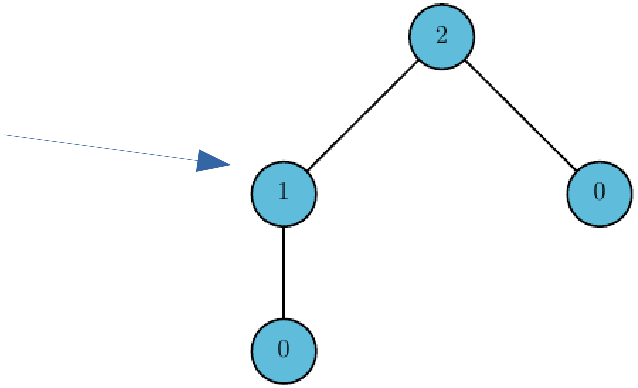
2 Typical accuracy definitions fail to account for data uncertainties, and complexity definition is largely arbitrary. The two are incommensurable.

$\Rightarrow$ Use information-theoretic *Minimum Description Length (MDL) principle*

Exhaustive Symbolic Regression

I. Function generation & optimisation

1) Generate all possible trees with given complexity = #nodes, with placeholder operators labelled by arity (number of arguments to operator)

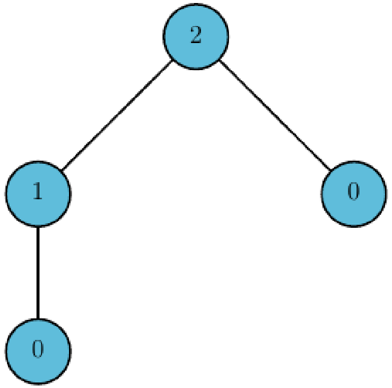


2) Decorate trees with all operator permutations

Exhaustive Symbolic Regression

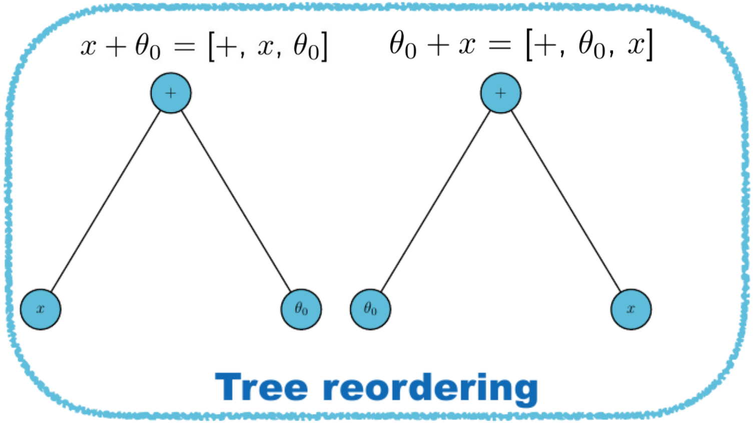
I. Function generation & optimisation

1) Generate all possible trees with given complexity = #nodes, with placeholder operators labelled by arity (number of arguments to operator)



2) Decorate trees with all operator permutations

3) Simplify and remove duplicates (*tree reordering*, *parameter permutations*, *simplifications*, *reparametrisation invariance*, *parameter combinations*)



4) Calculate maximum-likelihood parameter values

5) Repeat for all desired complexities

Exhaustive Symbolic Regression

II. Model selection principle: *minimum description length*

$$L(D) = L(H) + L(D \mid H)$$

Description length — Hypothesis — Residuals

- Purpose of functional fit is *data compression*
- Most information-efficient function has minimum *L(D)*

Exhaustive Symbolic Regression

II. Model selection principle: *minimum description length*

$$L(D) = L(H) + L(D|H)$$

Description length — Hypothesis — Residuals

$$L(D) = -\log(\mathscr{L}(\hat{\theta})) + k\log(n) + \sum_i^p \log(|\hat{\theta}_i|/\Delta_i) + p\log(2) + \sum_j \log(c_j)$$

Shannon-Fano coding

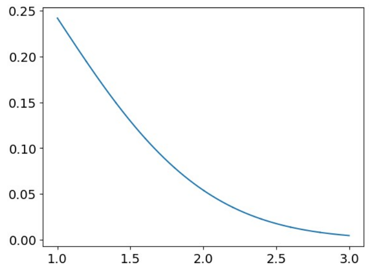
k nodes with n basis functions needs $\log\left(n^k\right)$ nats

Send p parameters, each with precision Δ_i

Integer constants

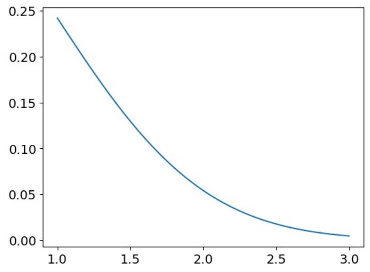
- Purpose of functional fit is *data compression*
- Most information-efficient function has minimum *L(D)*
- Both accuracy and complexity expressed in nats ⇒ can be combined
- Accounts for both functional and parametric complexity. Accuracy is likelihood.

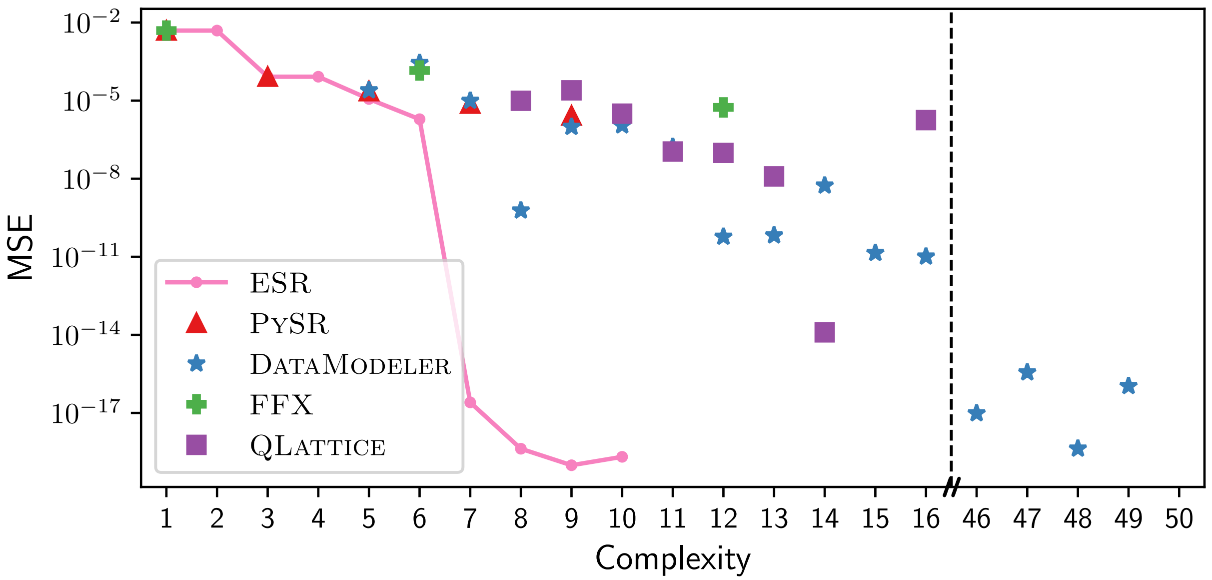
Test case 0: Benchmarking



- *feynman_I_6_2a* dataset from the *SRBench 2022 Competition*

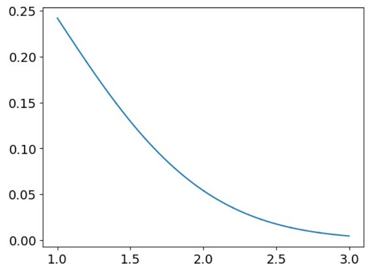
Test case 0: Benchmarking

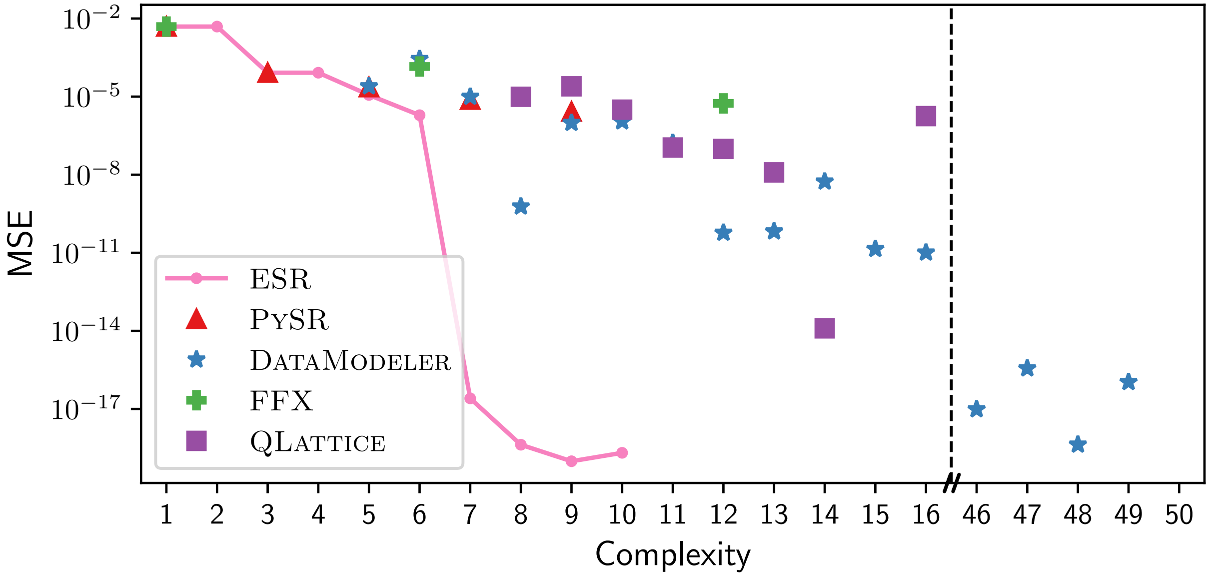


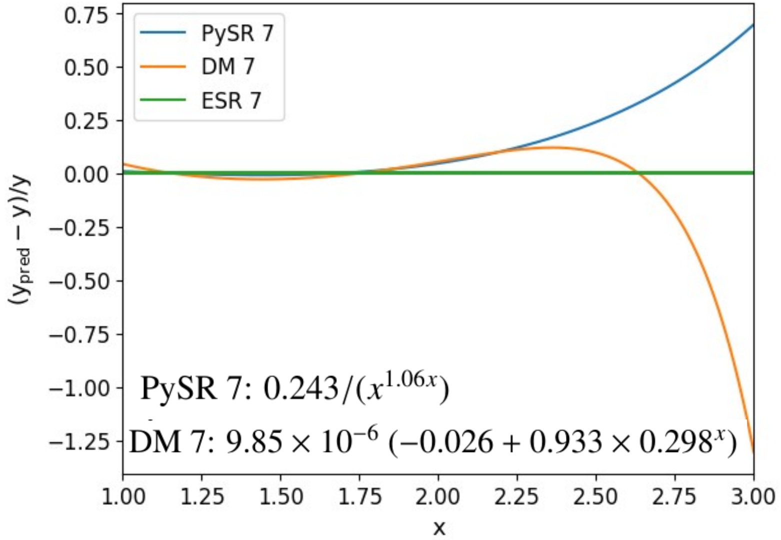


- *feynman_I_6_2a* dataset from the *SRBench 2022 Competition*

Test case 0: Benchmarking



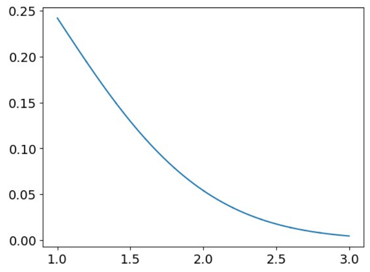


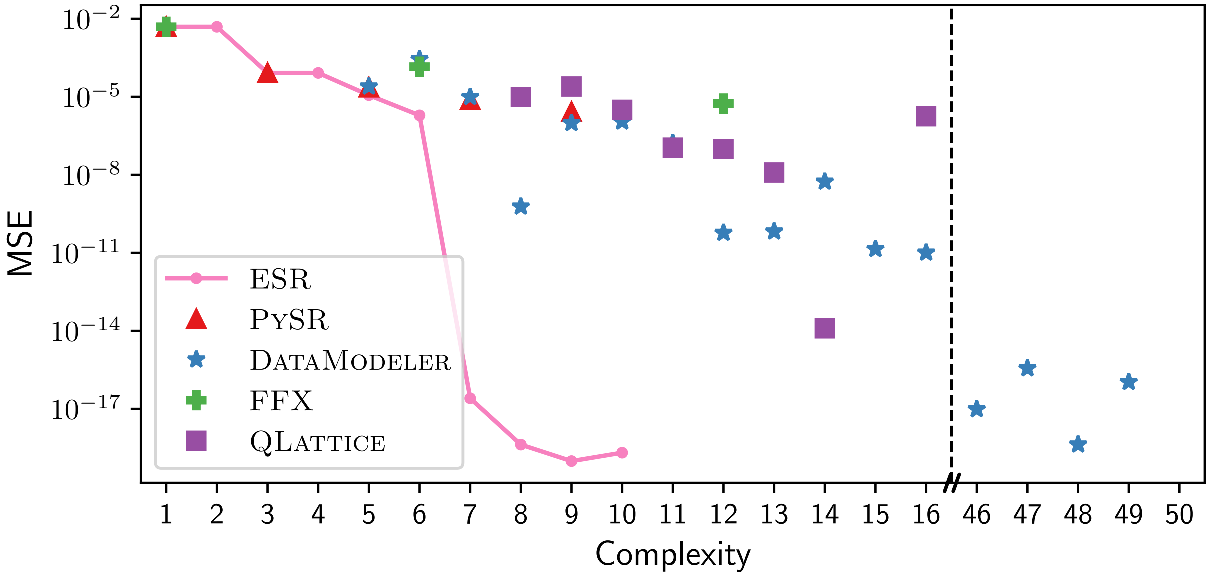


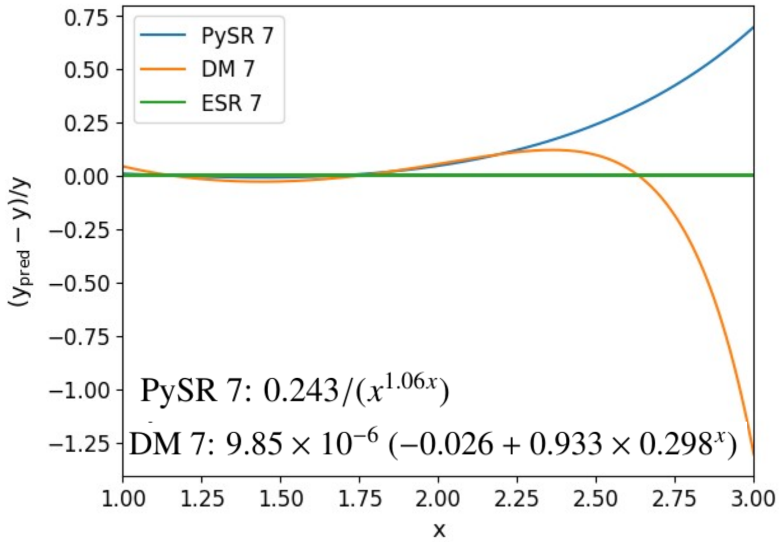
- *feynman_I_6_2a* dataset from the *SRBench 2022 Competition*
- Not only does ESR get by far the lowest error...

$$y = \theta_1 \theta_0^{x^2}$$
$$\theta_0 = 0.6065$$
$$\theta_1 = 0.3989$$

Test case 0: Benchmarking







$$y = \theta_1 \theta_0^{x^2}$$

$$\theta_0 = 1/\sqrt{e}$$

$$\theta_1 = 1/\sqrt{2\pi}$$

- *feynman_I_6_2a* dataset from the *SRBench 2022 Competition*
- Not only does ESR get by far the lowest error... it discovers the standard normal!

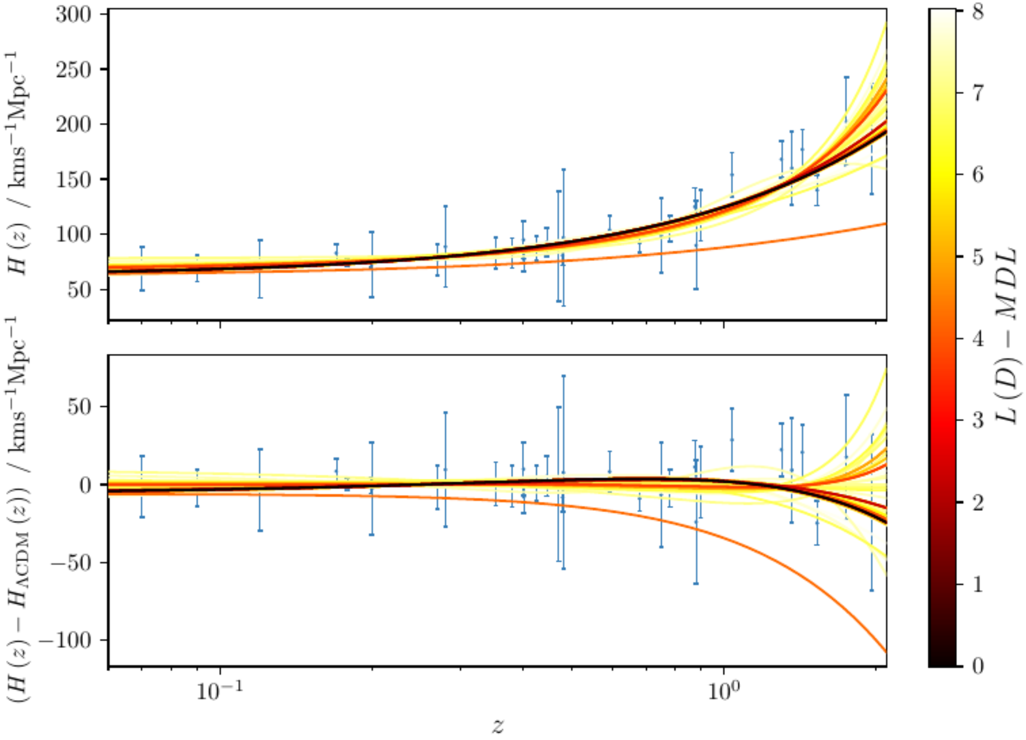
Test case 1: The law of cosmic expansion

- Can we determine the functional form of cosmic expansion without assuming GR?
- How good is the Friedmann equation relative to other simple functions?

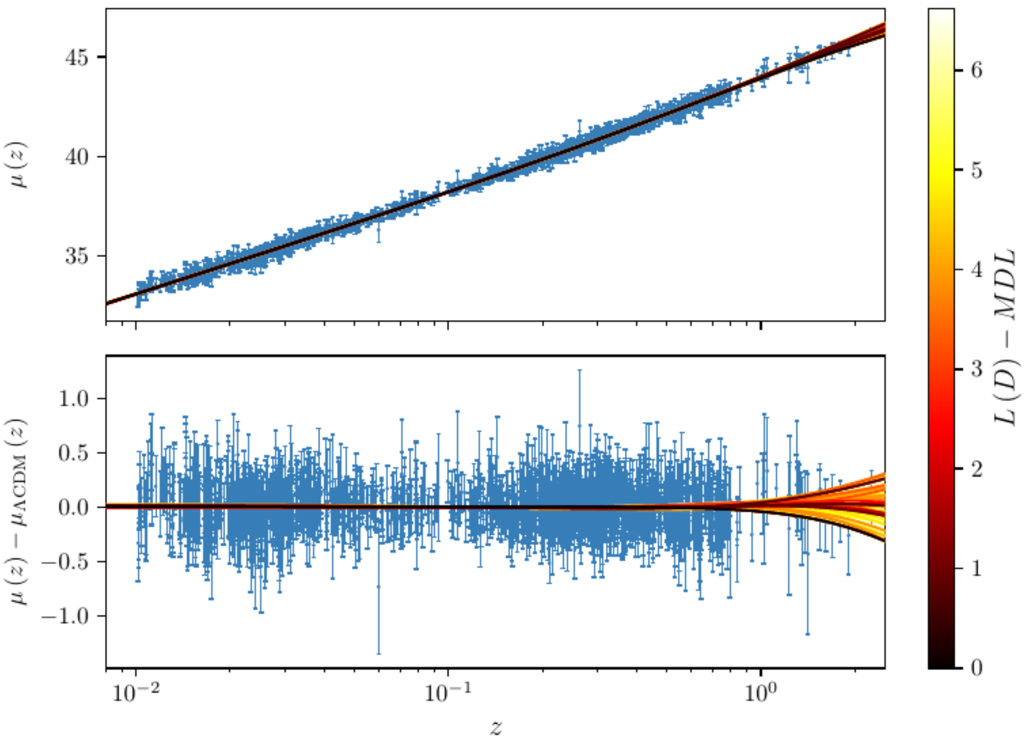
$$H(z)^2_{\Lambda\mathrm{CDM}} = \theta_0 + \theta_1\,(1+z)^3 \qquad H(z)^2_{\Lambda\mathrm{fluid}} = \theta_0 + \theta_1\,(1+z)^{\theta_2}$$

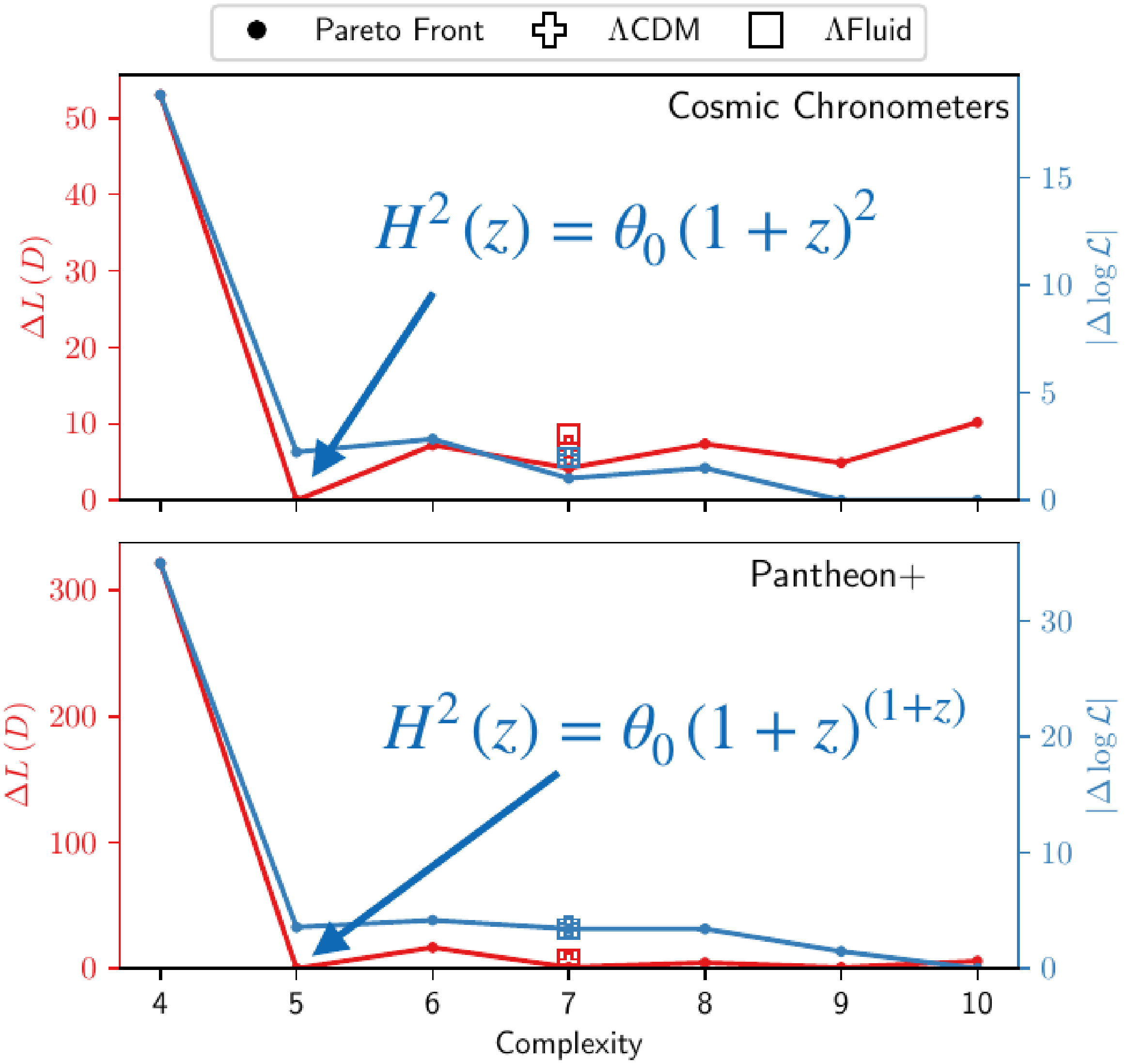
- Data:
  - Cosmic chronometers (32 data points) (Moresco et al 2022)
  - Type Ia Supernovae (1590 data points) (Pantheon+, Scolnic et al 2021)
- Basis operators: $\{x \equiv 1+z, \theta, \mathrm{inv}, +, -, \times, \div, \mathrm{pow}\}$

Cosmic Chronometers



Pantheon+ Supernovae





- ΛCDM ranked 39th for cosmic chronometers and 37th for SNe
- Best functions approximate ΛCDM at low z, but are simpler
- ~200 functions (up to complexity 10) more accurate than ΛCDM for Pantheon+

Test case 2: Potential of the inflaton

What we know

$$A_S = (0.027 \pm 0.0027)\, M_{\rm Pl},$$
$$n_S = 0.9649 \pm 0.0042,$$
$$r < 0.028 \;\; (95\%\, \mathrm{CL}).$$

Test case 2: Potential of the inflaton

What we know

$$A_S = (0.027 \pm 0.0027)\, M_{\rm Pl},$$
$$n_S = 0.9649 \pm 0.0042,$$
$$r < 0.028 \ (95\% \ \text{CL}).$$

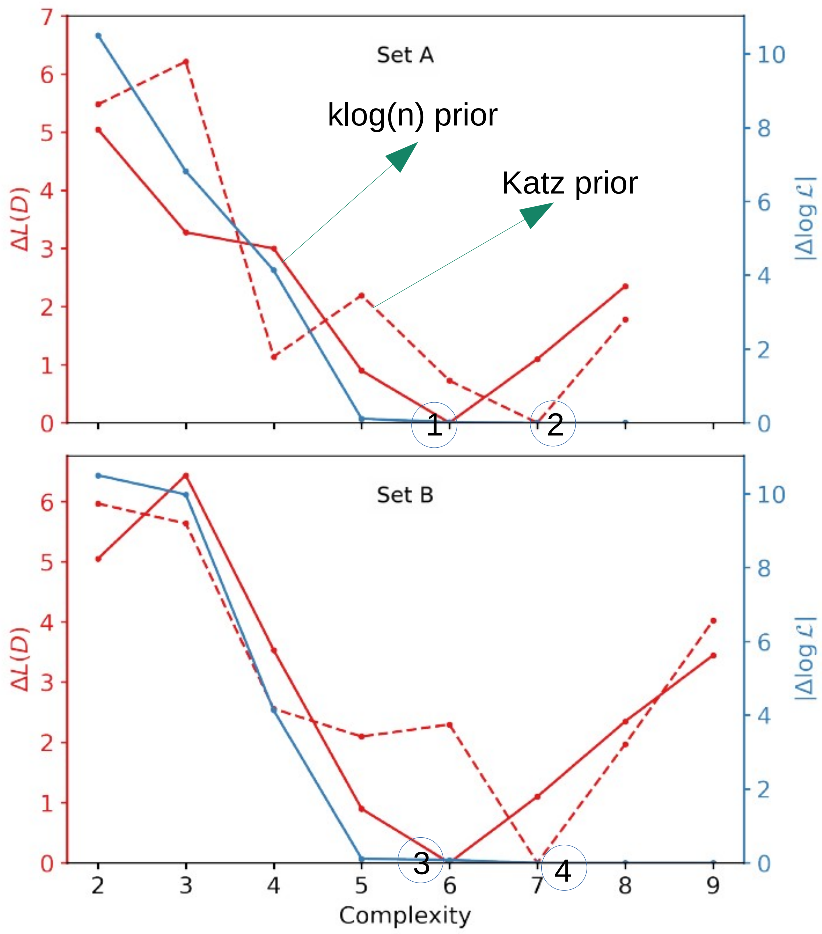
Operator basis sets

A: $\left\{x, a, \text{inv}, \exp, \log, \sin, \sqrt{|.|}, \text{cube}, \text{square}, +, *, -, /\right\}$

B: $\left\{x, a, \text{inv}, \exp, \log, +, *, -, /, \text{power}\right\}$

Best functions

1. $e^{-e^{e^{e^{\phi}}}}$
2. $\theta_0 \left(\theta_1 + \log(\phi)^2\right)$
3. $e^{-e^{e^{e^{\phi}}}}$
4. $\theta_0 \phi^{\frac{\theta_1}{\phi}}$



Operator set B, klog(n) prior

Rank	$V(\phi)$	Complexity	Prediction		Description length			
			n_s	r	Residuals	Function	Parameter	Total
1	$e^{-e^{e^{e^{\phi}}}}$	6	0.9678	0.002	-12.65	6.59	0.00	-6.06
2	$\theta_0 e^{-e^{\phi}}$	5	0.9668	0.002	-12.79	6.93	0.69	-5.16
3	$\lvert\theta_0\rvert^{e^{e^{\phi}}}$	5	0.9674	0.002	-12.72	6.93	0.69	-5.09
4	$e^{e^{\phi}-e^{e^{e^{\phi}}}}$	8	0.9685	0.002	-12.50	8.79	0.00	-3.71
5	$\theta_0 e^{-e^{e^{\phi}}}$	6	0.9680	0.002	-12.62	8.32	0.69	-3.61
6	$e^{-e^{\frac{1}{\phi}}}$	5	0.9768	0.008	-8.65	5.49	0.00	-3.16
7	$e^{\theta_0 e^{e^{\phi}}}$	6	0.9674	0.002	-12.72	8.32	1.36	-3.04
8	$\lvert\theta_0\rvert^{e^{e^{e^{\phi}}}}$	6	0.9678	0.002	-12.65	8.32	1.39	-2.94
9	$e^{e^{\phi}} e^{-e^{e^{e^{\phi}}}}$	9	0.9685	0.002	-12.50	9.89	0.00	-2.62
10	$\lvert\theta_0\rvert^{\frac{1}{\phi}}$	4	0.9756	0.019	-8.77	5.55	0.69	-2.53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1272	$\theta_0\left(1-e^{-\sqrt{2/3}\phi}\right)^2$	9	0.9678	0.003	-12.63	17.51	0.69	5.57
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8697	$\theta_0\phi^2$	4	0.9669	0.132	31.82	5.55	0.69	38.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10839	$\theta_0\phi^4$	5	0.9508	0.262	168.23	6.93	0.69	178.81

Conclusions

- *Exhaustive Symbolic Regression*: Guaranteed to find best simple function for any data
- *Minimum description length* affords principled combination of accuracy and simplicity
- A *Katz language model* can assign function priors based on a training set

https://github.com/DeaglanBartlett/ESR https://github.com/DeaglanBartlett/Katz

Conclusions

- *Exhaustive Symbolic Regression*: Guaranteed to find best simple function for any data
- *Minimum description length* affords principled combination of accuracy and simplicity
- A *Katz language model* can assign function priors based on a training set

https://github.com/DeaglanBartlett/ESR https://github.com/DeaglanBartlett/Katz

So far we've learnt...

1 Cosmic chronometers and SNe don't uniquely favour Friedmann

2 Planck implies optimal inflationary potentials

3 The radial acceleration relation doesn't clearly support modified gravity

Extra Slides

MDL as a Bayesian statistic

$$P\left(f_i|D\right) = \frac{1}{P(D)} \int P\left(D|f_i, \boldsymbol{\theta}_i\right) P\left(\boldsymbol{\theta}_i|f_i\right) P\left(f_i\right) \mathrm{d}\boldsymbol{\theta}_i$$

$$\log P\left(f_i|D\right) = -\log P\left(f_i\right) - \log \mathcal{Z}\left(D|f_i\right)$$

$$\log \mathcal{Z}\left(D|f_i\right) \simeq \log H(D, f_i, \hat{\boldsymbol{\theta}}_i) + \frac{p}{2}\log 2\pi - \frac{1}{2}\log\left|\det \hat{\boldsymbol{I}}^H\right|$$

$$\left(\hat{I}^H_{\alpha\beta} = -\partial_\alpha \partial_\beta \log H(D, f_i, \boldsymbol{\theta}_i)|_{\hat{\boldsymbol{\theta}}_i} \right)$$

Functional Priors

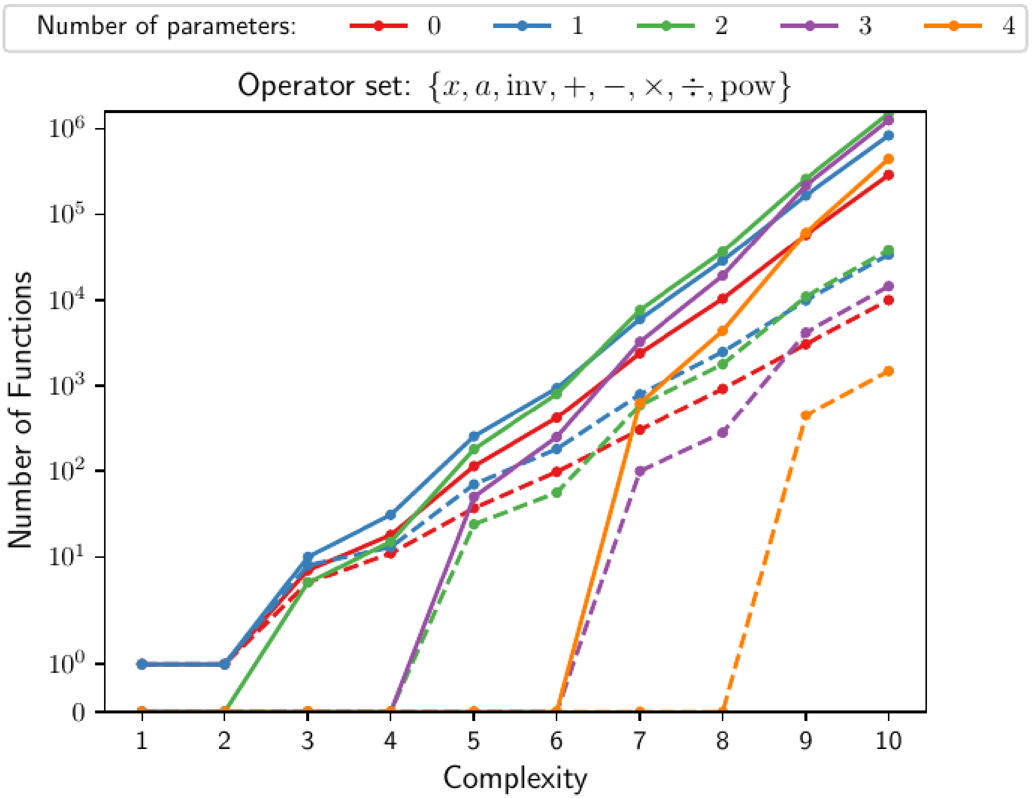
MDL implies $-\log P\left(f_i\right) = k\log n + \sum_\alpha \log c_\alpha$

OR

Quantify our lower prior on e.g. *sin(sin(x_0+x_1))* compared to *sin(x_0)+sin(x_1)*

"Katz back-off model" determines probability of next operator given *n* preceding operators based on a training set of equations

Simplifications make an exhaustive search feasible



168 million

Naïve estimate: $\sum_{j=1}^{n} j^k = H_n^{(-k)}$

5.2 million

Total number of decorated trees

134,234

Number of unique equations

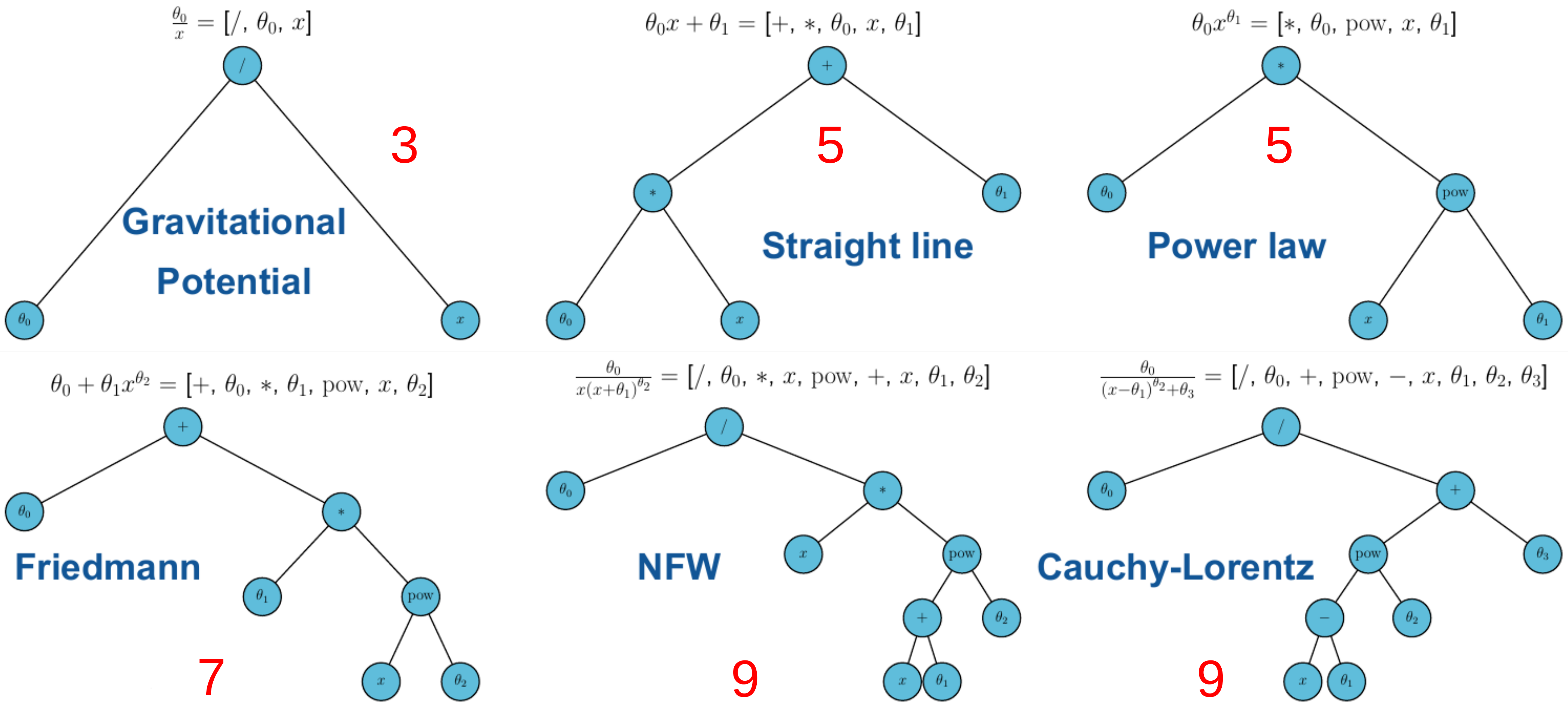
119,861

Number of equations containing at least one parameter

1400

Times fewer equations to consider than our naïve guess

Many physics functions have complexity < 10



Precision of constants - tradeoff between accuracy and information needed

True $\hat{\theta}_i$ uniformly distributed within $\pm\Delta_i/2$ of transmitted value - may not transmit true MLP

Taylor expand log-likelihood about true value:

$$-\log(\mathscr{L}(\hat{\theta}+\mathbf{d})) \approx -\log(\mathscr{L}(\hat{\theta})) + \frac{1}{2}\mathbf{d}^{\mathrm{T}}\,\mathrm{I}\,\mathbf{d} \qquad \mathrm{I}_{ij} = \left.\frac{\mathrm{d}^2(-\log\mathcal{L})}{\mathrm{d}\theta_i\,\mathrm{d}\theta_j}\right|_{\hat{\theta}}$$

Giving expected contribution to description length

$$L(\boldsymbol{\Delta}) = \frac{1}{2}\sum_{ij}\langle \mathrm{I}_{ij} d_i d_j \rangle - \sum_i \log(\Delta_i)$$

Minimise this:

$$L(\Delta_i) = \frac{1}{24}\mathrm{I}_{ii}\Delta_i^2 - \log(\Delta_i) \quad \Longrightarrow \quad \Delta_i = \left(\frac{12}{\mathrm{I}_{ii}}\right)^{1/2}$$

The Description Length of a Function

$$L(D) = -\log(\mathscr{L}(\hat{\theta})) + k\log(n) - \frac{p}{2}\log(3) + \sum_{i}^{p}\left(\frac{1}{2}\log(\mathrm{I}_{ii}) + \log(|\hat{\theta}_i|)\right)$$

$$\frac{1}{2}\left(p\log(N) - 2\log(\mathscr{L})\right) = \frac{1}{2}\mathrm{BIC}$$

(for large number of data points, N)

- Description length looks like BIC plus corrections due to structural complexity (prior on model)
- For large N, equivalent to minimising the BIC (an approximation to the evidence)

$$P(H) = \exp(-L(D)) / \sum(\exp(-L(D)))$$

Cosmic Chronometers - Standard Clocks

$$H(z) = -\frac{1}{1+z}\frac{\mathrm{d}z}{\mathrm{d}t} \approx -\frac{1}{1+z}\frac{\delta z}{\delta t}$$

Image credit: AAS NOVA

Jimenez & Loeb 2001 (arXiv:astro-ph/0106145)
Moresco 2015 (arXiv:1503.01116)
Moresco et al . 2016 (arXiv:1601.01701)
Ratsimbazafy et al. 2017 (arXiv:1702.00418)
Stern et al. 2010 (arXiv:0907.3149)
Simon et al. 2004 (arXiv:astro-ph/0412269)
Borghi et al. 2021 (arXiv:2110.04304)
Zhang et al. 2012 (arXiv:1207.4541)
Moresco et al. 2012 (arXiv:1201.3609)
Moresco et al. 2022 (arXiv:2201.07241)

- Passively evolving stellar population are standard clocks - measure δt
- Directly measure δz
- (Cosmological) model-independent measurement of $H(z)$
- We use sample of 32 CC $H(z)$ measurements

Type Ia Supernovae - Standard Candles

$$d_{\mathrm{L}}(z) = (1+z)\int_0^z \frac{\mathrm{d}z'}{H(z')}$$

$$\mu(z) = 5\log_{10}\left(\frac{d_{\mathrm{L}}(z)}{10\,\mathrm{pc}}\right)$$

$$\mu = m_{\mathrm{B}} + \alpha x_1 - \beta c - M_0$$

Amplitude (m_{B}), **Stretch** (αx_1), **Colour** (βc), **Rest-frame magnitude** (M_0)

Pantheon+ sample with SH0ES Cepheid calibration

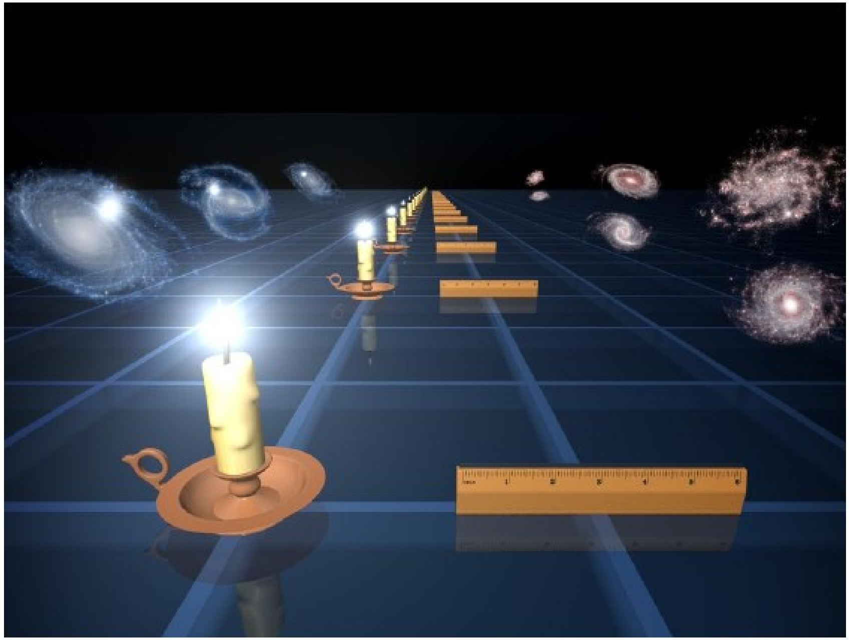


Image credit: NASA

Scolnic et al. 2021 (arXiv:2112.03863)
Riess et al. 2022 (arXiv:2112.04510)

CCs

Rank	$y(x)$ / $\mathrm{km^2 s^{-2} Mpc^{-2}}$	Complexity	Parameters			Codelength			
			θ_0	θ_1	θ_2	Residuals[1]	Function[2]	Parameter[3]	Total
1	$\theta_0 x^2$	5	3883.44	-	-	8.36	5.49	2.53	16.39
2	$\lvert\theta_0\rvert^{x^{\theta_1}}$	5	3982.43	0.22	-	7.97	5.49	5.24	18.70
3	$\theta_0 \lvert\theta_1\rvert^{-x}$	5	1414.43	0.31	-	7.57	6.93	5.58	20.08
4	$\theta_0 x^{\theta_1}$	5	3834.51	2.03	-	8.35	6.93	5.08	20.36
5	$x^2(\theta_0 + x)$	7	3881.85	-	-	8.36	9.70	2.53	20.60
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
39	$\theta_0 + \theta_1 x^3$	9	3164.02	1481.71	-	7.28	12.48	3.76	23.51
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
84	$\theta_0 + \theta_1 x^{\theta_2}$	7	3322.96	1374.97	3.08	7.27	11.27	6.52	25.06

[1] $-\log \mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n) + \sum_j \log(c_j)$ [3] $-\frac{p}{2}\log(3) + \sum_i^p\left(\frac{1}{2}\log(I_{ii}) + \log(\lvert\hat{\theta}_i\rvert)\right)$

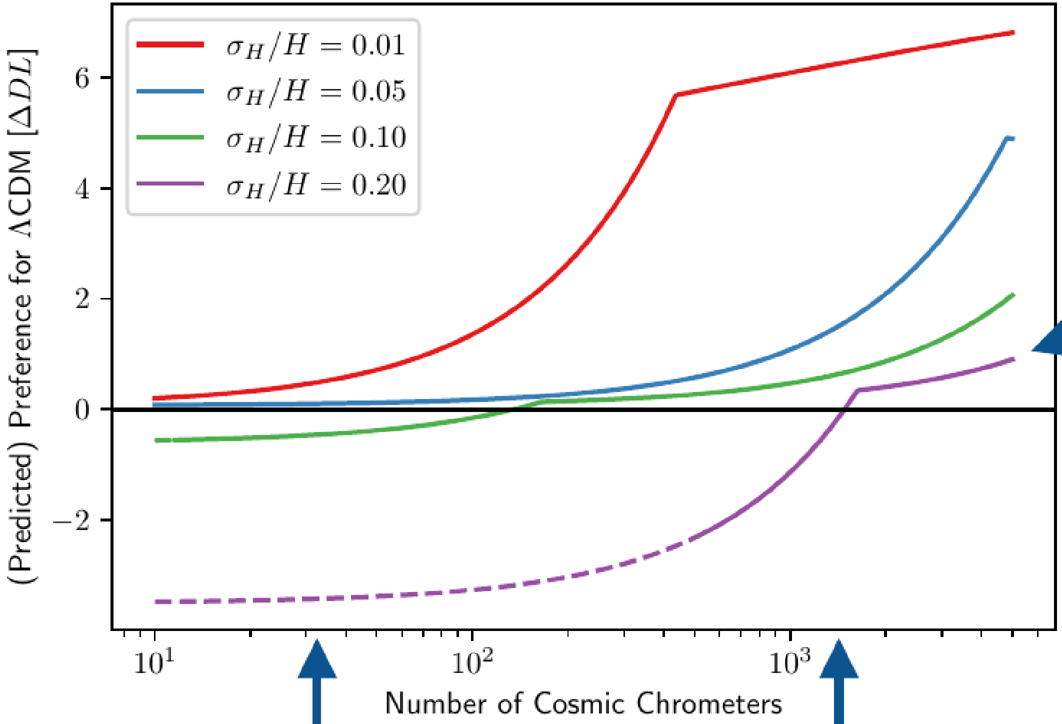
SNe

Rank	$y(x)$ / $\mathrm{km^2 s^{-2} Mpc^{-2}}$	Complexity	Parameters			Codelength			
			θ_0	θ_1	θ_2	Residuals[1]	Function[2]	Parameter[3]	Total
1	$\theta_0 x^x$	5	5345.02	-	-	706.18	6.93	5.11	718.22
2	$\lvert\theta_0\rvert^{x^{\theta_1}}$	9	5280.11	0.16	-	705.11	5.49	8.41	719.01
3	$\theta_0 \lvert\theta_1\rvert^{-x}$	5	1694.95	0.32	-	701.79	6.93	10.33	719.05
4	$\theta_0 x^{x^{\theta_1}}$	7	5378.69	0.78	-	702.45	9.70	6.98	719.13
5	$\lvert\theta_0\rvert^{\lvert\theta_1\rvert^x}$	5	1898.47	1.14	-	701.88	5.49	12.64	720.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
37	$\theta_0 + \theta_1 x^3$	9	3591.09	1773.63	-	701.85	12.48	8.81	723.13
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	$\theta_0 + \theta_1 x^{\theta_2}$	7	3280.83	2069.32	2.73	701.64	11.27	12.19	725.10

[1] $-\log \mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n) + \sum_j \log(c_j)$ [3] $-\frac{p}{2}\log(3) + \sum_i^p\left(\frac{1}{2}\log(I_{ii}) + \log(\lvert\hat{\theta}_i\rvert)\right)$

Should we have seen ΛCDM? No.

Mock cosmic chronometer data assuming ΛCDM (Planck18)



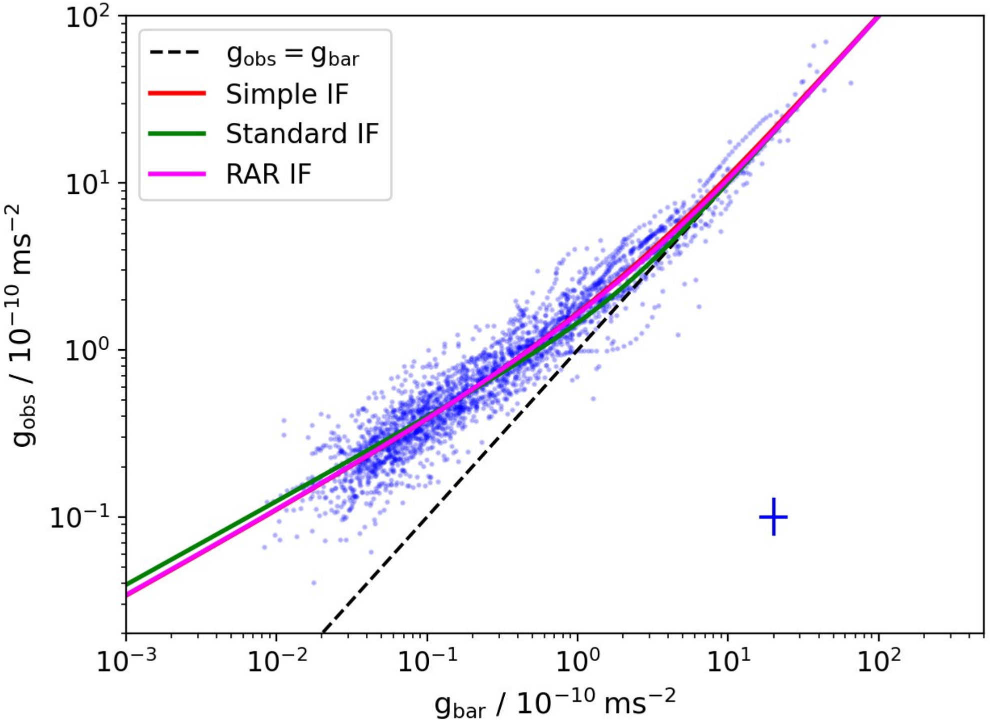
Current observational uncertainties

We used this many

Would need this many

Test case 2: The radial acceleration relation

- Relates acceleration sourced by baryons (g_{bar}) to total acceleration as measured by rotation velocity (g_{obs})
- 2,696 points from 153 late-type galaxies (SPARC sample)
- Regularity and low scatter hard to understand in ΛCDM



<u>MOND Interpolating Functions (IFs)</u>

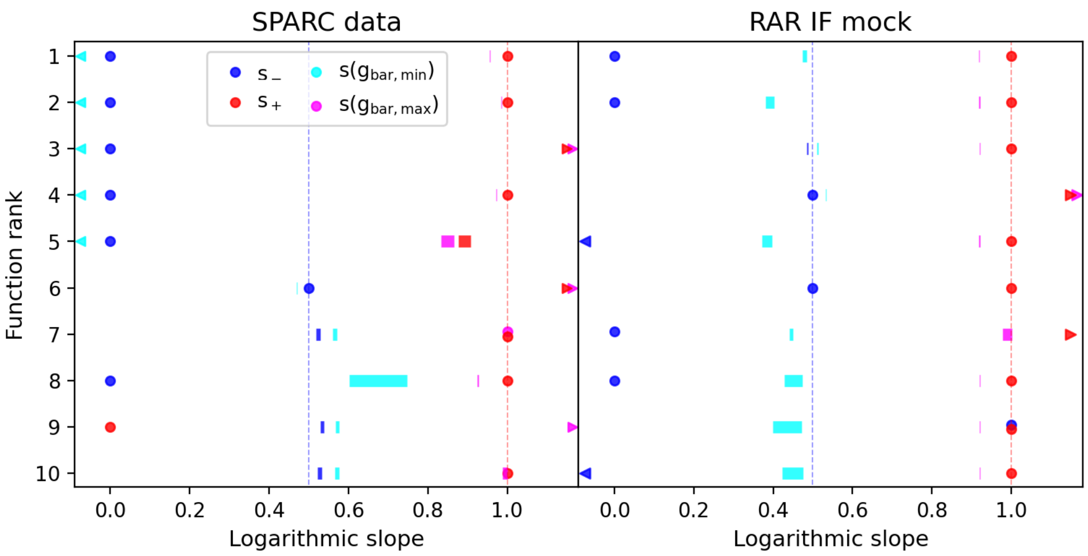
Simple — $g_{obs} = g_{bar}/2 + \sqrt{g_{bar}^2/4 + g_{bar}\, a_0}$

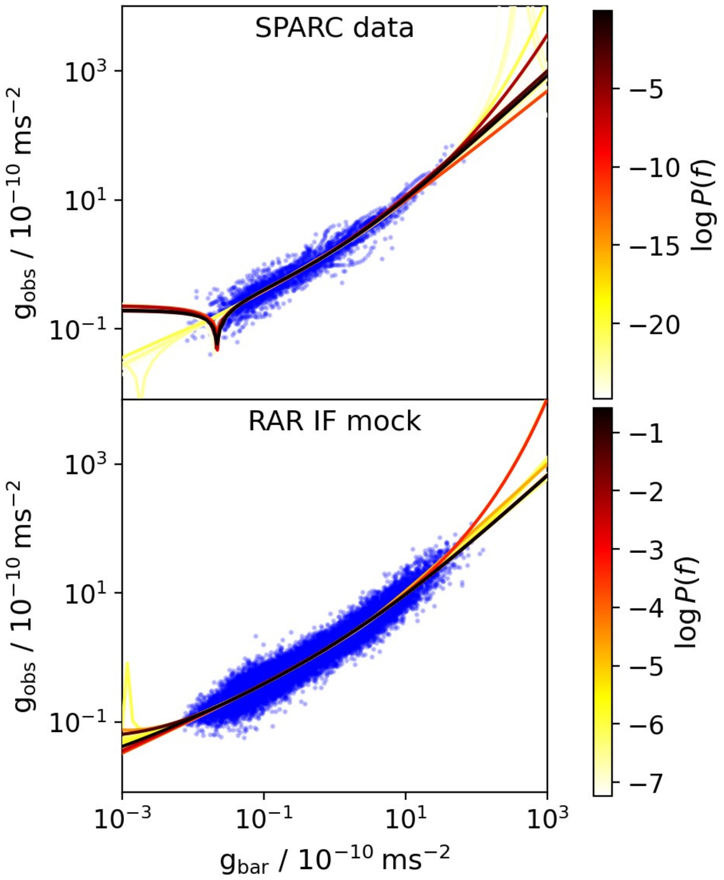
Standard — $g_{obs} = \frac{1}{\sqrt{2}}\sqrt{g_{bar}^2 + \sqrt{g_{bar}^2\left(g_{bar}^2 + 4g_0^2\right)}}$

RAR — $g_{obs} = g_{bar}/(1 - \exp(-\sqrt{g_{bar}/g_0}))$

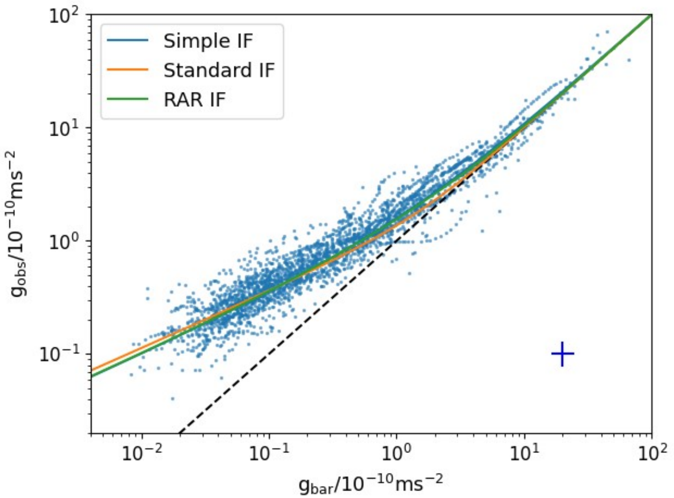
1) Are the MOND IFs optimal descriptions of the RAR?

2) Do optimal solutions satisfy the MOND limits (and hence may be considered new IFs)?

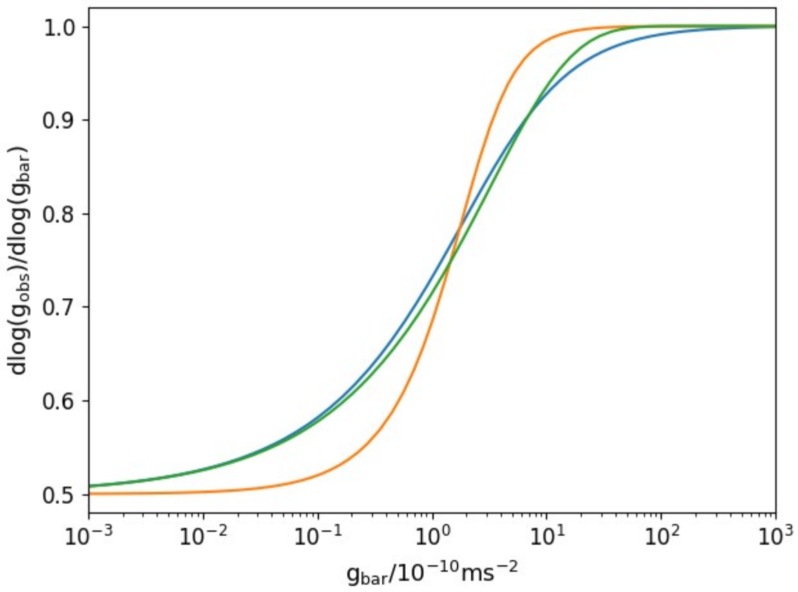


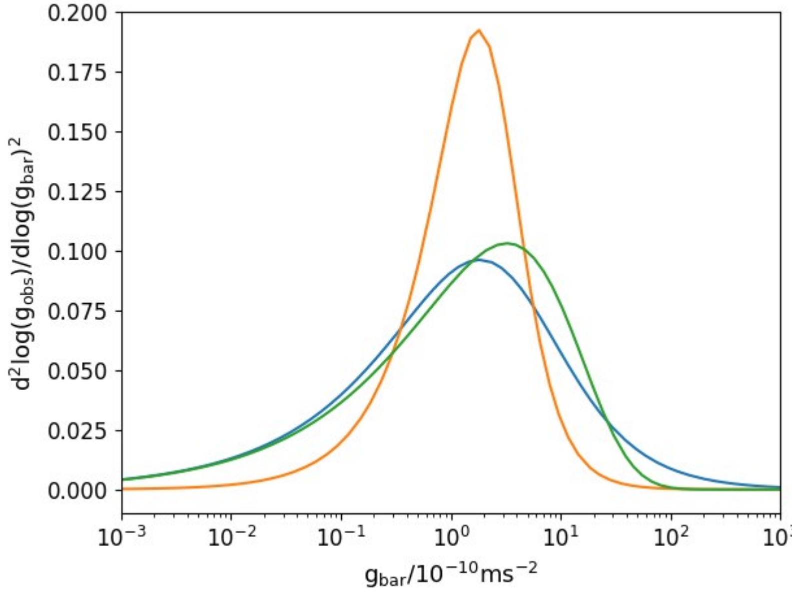


- Newtonian limit often found; deep-MOND limit rarely
- Can't recover MOND behaviour even from MOND mocks!
  $\Rightarrow$ Precision and dynamic range of data insufficient

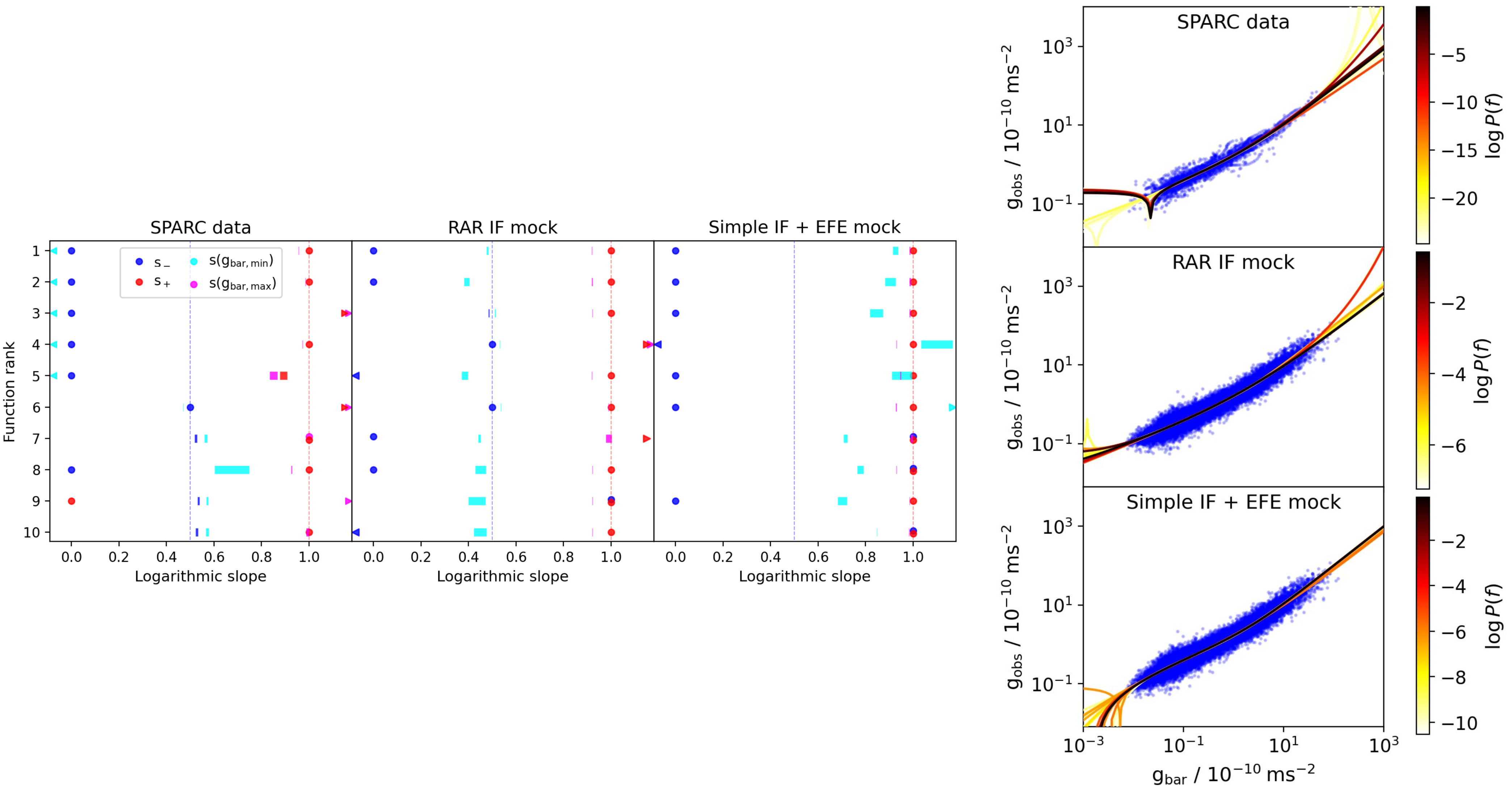


First logarithmic derivative





Second logarithmic derivative



SPARC data
RAR IF mock
Simple IF + EFE mock
Function rank
Logarithmic slope
s_-
s_+
$s(g_{bar,min})$
$s(g_{bar,max})$
g_{obs} / 10^{-10} ms^{-2}
g_{bar} / 10^{-10} ms^{-2}
$\log P(f)$

SPARC data

Rank	Function	Comp.	$P(f)$	Parameters				Description length			
				θ_0	θ_1	θ_2	θ_3	Resid.[1]	Func.[2]	Param.[3]	Total
1	$\theta_0\left(\lvert\theta_1+x\rvert^{\theta_2}+x\right)$	9	9.3×10^{-1}	0.84	-0.02	0.38	—	-1279.1	14.5	14.0	-1250.6
2	$\lvert\lvert\theta_1\rvert^x+\theta_0\rvert^{\theta_2}+x$	9	6.4×10^{-2}	-0.99	0.64	0.36	—	-1279.9	12.5	19.6	-1247.9
3	$\lvert\theta_0\rvert^{\lvert\theta_1-x\rvert^{\theta_2}-\theta_3}$	9	2.0×10^{-3}	-1.4×10^{2}	0.02	0.14	0.89	-1276.4	12.5	19.5	-1244.4
4	$\lvert\theta_0(\theta_1+x)\rvert^{\theta_2}+x$	9	1.4×10^{-4}	0.35	-0.02	0.34	—	-1268.9	14.5	12.7	-1241.7
5	$\left\lvert\theta_0-\lvert\theta_1-x\rvert^{\theta_2}\right\rvert^{\theta_3}$	9	1.0×10^{-5}	-0.30	0.02	0.42	2.14	-1271.1	12.5	19.5	-1239.1
6	$\sqrt{x}\exp\left(\frac{\lvert\theta_0+x\rvert^{\theta_1}}{2}\right)$	9	1.5×10^{-9}	-0.02	0.36	—	—	-1257.9	17.5	10.0	-1230.3
7	$\left(\frac{\lvert\theta_0\rvert^x}{x}\right)^{\theta_1}+x$	9	2.4×10^{-10}	1.87	-0.52	—	—	-1250.6	14.5	7.6	-1228.5
8	$\sqrt{\lvert\theta_0+x\rvert}+\theta_1 x$	8	1.8×10^{-10}	-1.8×10^{-3}	0.72	—	—	-1245.6	12.9	4.5	-1228.2
9	$\left\lvert\theta_0+\frac{1}{\sqrt[4]{x}}\right\rvert^{\theta_1}$	8	9.6×10^{-11}	-0.22	-2.14	—	—	-1251.1	14.3	9.2	-1227.6
10	$\left(\sqrt{x}+\frac{1}{x}\right)^{\theta_0}+x$	9	8.2×10^{-11}	-0.53	—	—	—	-1248.3	16.1	4.8	-1227.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	$x/(\exp(\theta_0)-\lvert\theta_1\rvert^{\sqrt{x}})$	9	2.2×10^{-11}	0.03	0.44	—	—	-1250.9	17.5	7.3	-1226.1
—	Double power law	11	9.7×10^{-16}	4.65	3.96	1.03	0.60	-1252.3	17.7	18.5	-1216.1
—	Simple IF	10	5.5×10^{-25}	1.11	—	—	—	-1217.3	18.6	3.9	-1194.8
—	RAR IF	9	6.7×10^{-26}	1.13	—	—	—	-1212.8	16.1	3.9	-1192.7
—	Simple IF + EFE	59	5.0×10^{-69}	1.16	6.8×10^{-3}	—	—	-1238.9	139.9	5.6	-1093.4
—	Standard IF	14	9×10^{-150}	1.54	—	—	—	-939.5	27.9	4.1	-907.5

[1] $-\log\mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n)+\sum_j\log(c_j)$ [3] $-\frac{p}{2}\log(3)+\sum_i^p(\log(I_{ii})^{1/2}+\log(\lvert\hat{\theta}_i\rvert))$

RAR IF mock

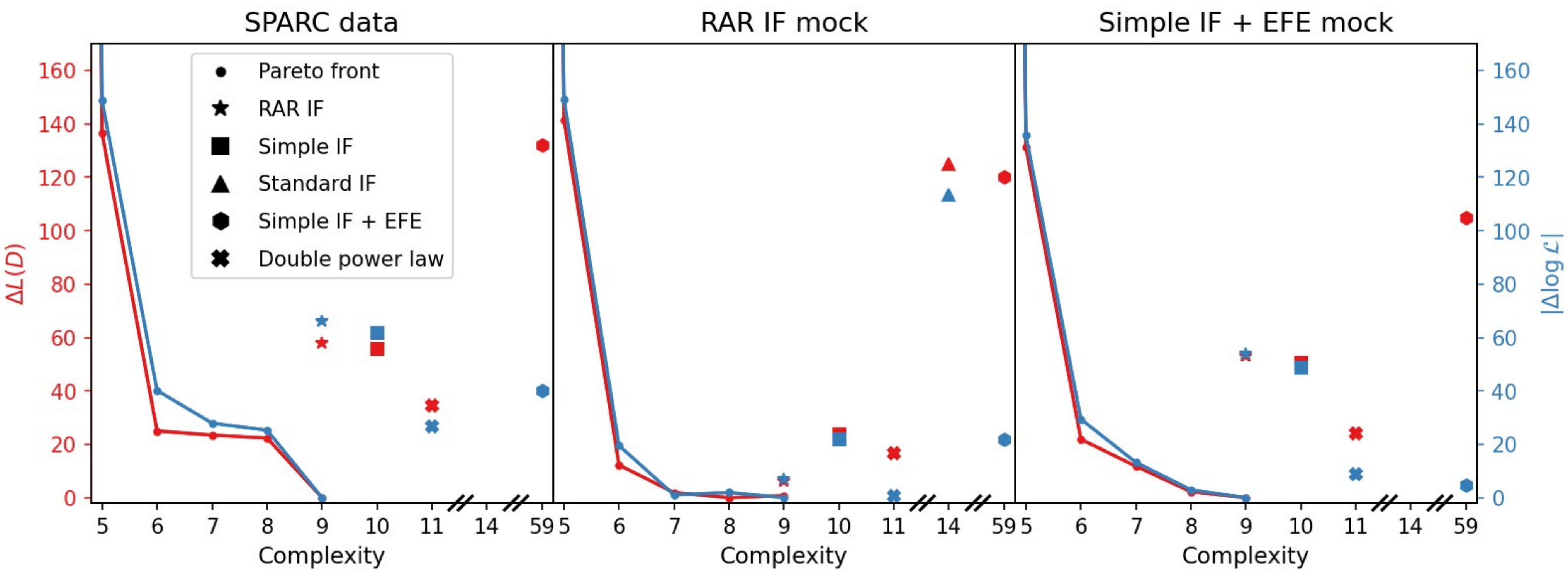
Rank	Function	Comp.	$P(f)$	Parameters				Description length			
				θ_0	θ_1	θ_2	θ_3	Resid.[1]	Func.[2]	Param.[3]	Total
1	$\theta_0 + \theta_1 x + \sqrt{x}$	8	5.6×10^{-1}	9.1×10^{-3}	0.63	—	—	-2045.2	12.9	4.9	-2027.4
2	$\sqrt{\lvert\theta_0 + x\rvert} + \theta_1 x$	8	2.8×10^{-1}	3.0×10^{-3}	0.64	—	—	-2044.4	12.9	4.8	-2026.7
3	$\theta_0 x + x^{\theta_1}$	7	8.2×10^{-2}	0.64	0.49	—	—	-2045.2	11.3	8.5	-2025.5
4	$\sqrt{x}\exp\left(\frac{x^{\theta_0}}{2}\right)$	7	3.5×10^{-2}	0.36	—	—	—	-2040.7	12.5	3.5	-2024.7
5	$(\theta_0 + x)\left(\theta_1 + \frac{1}{\sqrt{x}}\right)$	9	1.1×10^{-2}	1.3×10^{-3}	0.64	—	—	-2044.5	16.1	4.8	-2023.5
6	$\frac{1}{\sqrt{\lvert\theta_0 + \frac{1}{x}\rvert}} + x$	8	8.8×10^{-3}	1.74	—	—	—	-2038.5	12.9	2.3	-2023.3
7	$(x\lvert\theta_0\rvert)^{(x\lvert\theta_1\rvert)^{\theta_2}}$	9	3.1×10^{-3}	-2.09	-1.4×10^{-4}	0.04	—	-2045.3	12.5	10.6	-2022.2
8	$\theta_0 x + \lvert\theta_1 + x\rvert^{\theta_2}$	9	2.4×10^{-3}	0.64	1.4×10^{-3}	0.49	—	-2045.4	14.5	8.9	-2022.0
9	$x\left(\lvert\theta_0 - x\rvert^{\theta_1} - \theta_2\right)$	9	2.3×10^{-3}	1.2×10^{-3}	-0.51	-0.64	—	-2045.3	14.5	8.9	-2021.9
10	$(\theta_0 - x)\left(\theta_1 - x^{\theta_2}\right)$	9	2.2×10^{-3}	-6.5×10^{-4}	-0.64	-0.51	—	-2045.4	14.5	9.0	-2021.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
27	$x/(\exp(\theta_0) - \exp(-\sqrt{x}))$	9	3.2×10^{-4}	-0.01	—	—	—	-2039.3	17.5	1.9	-2020.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	$x/(\exp(\theta_0) - \lvert\theta_1\rvert^{\sqrt{x}})$	9	1.1×10^{-4}	-5.0×10^{-3}	0.38	—	—	-2042.1	17.5	5.7	-2018.9
—	RAR IF	9	1.0×10^{-3}	1.14	—	—	—	-2041.1	16.1	3.9	-2021.1
—	Double power law	11	3.4×10^{-8}	1.25	1.47	0.90	0.54	-2047.2	17.7	18.7	-2010.8
—	Simple IF	10	2.8×10^{-11}	1.12	—	—	—	-2026.2	18.6	3.9	-2003.7
—	Standard IF	14	2.9×10^{-55}	1.54	—	—	—	-1934.4	27.9	4.1	-1902.4
—	Simple IF + EFE	59	5.9×10^{-64}	1.12	0	—	—	-2026.2	139.9	3.9	-1882.4

[1] $-\log\mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n) + \sum_j \log(c_j)$ [3] $-\frac{p}{2}\log(3) + \sum_i^p (\log(I_{ii})^{1/2} + \log(\lvert\hat{\theta}_i\rvert))$

Simple IF + EFE mock

Rank	Function	Comp.	$P(f)$	Parameters θ_0	θ_1	θ_2	θ_3	Description length Resid.[1]	Func.[2]	Param.[3]	Total
1	$\theta_0 + \sqrt{x^2+2x}$	9	8.9×10^{-1}	-0.06	—	—	—	-2017.7	14.5	3.1	-2000.0
2	$\theta_0 + \sqrt{x\lvert\theta_1+x\rvert}$	8	9.3×10^{-2}	-0.06	1.97	—	—	-2017.9	12.9	7.3	-1997.8
3	$-\lvert\theta_0\rvert^{\sqrt{x}} + \theta_1 + x$	8	5.6×10^{-3}	0.26	0.95	—	—	-2017.9	12.9	10.1	-1995.0
4	$(\theta_0 - x)\left(\theta_1 - x^{\theta_2}\right)$	9	3.3×10^{-3}	3.1×10^{-3}	-0.71	-0.53	—	-2019.7	14.5	10.7	-1994.4
5	$x^{\theta_0} - \theta_1(\theta_2 - x)$	9	2.4×10^{-3}	0.39	0.79	0.12	—	-2020.9	14.5	12.3	-1994.1
6	$\lvert\theta_0 - x\rvert^{\theta_1} - \theta_2 x$	9	2.0×10^{-3}	5.5×10^{-3}	0.48	-0.71	—	-2019.1	14.5	10.6	-1994.0
7	$x\lvert\theta_0\rvert^{-\lvert\theta_1\rvert^{x^{\theta_2}}}$	9	1.7×10^{-3}	0.04	-0.16	0.33	—	-2018.1	12.5	11.9	-1993.8
8	$x\left(\theta_0 + \lvert\theta_1 + x\rvert^{\theta_2}\right)$	9	1.5×10^{-3}	0.71	0.01	-0.53	—	-2018.7	14.5	10.6	-1993.7
9	$\lvert\theta_0\rvert^{\lvert\theta_1\rvert^{x^{\theta_2}}} + x$	9	6.5×10^{-4}	7.0×10^{-6}	0.03	0.17	—	-2016.7	12.5	11.4	-1992.8
10	$\exp\left(\theta_0 - \frac{1}{\sqrt[4]{x}}\right) + x$	9	5.5×10^{-4}	0.57	—	—	—	-2014.0	17.5	3.9	-1992.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	$x/(\exp(\theta_0) - \lvert\theta_1\rvert^{\sqrt{x}})$	9	1.8×10^{-5}	0.03	0.44	—	—	-2014.2	17.5	7.4	-1989.3
—	Double power law	11	3.4×10^{-11}	3.53	3.31	0.98	0.60	-2012.3	17.7	18.6	-1976.0
—	Simple IF	10	1.2×10^{-22}	1.11	—	—	—	-1972.1	18.6	3.9	-1949.6
—	RAR IF	9	7.0×10^{-24}	1.13	—	—	—	-1966.9	16.1	3.9	-1946.8
—	Simple IF + EFE	59	3.8×10^{-57}	1.19	8.6×10^{-3}	—	—	-2016.0	139.9	5.9	-1870.2
—	Standard IF	14	2×10^{-141}	1.54	—	—	—	-1708.3	27.9	4.1	-1676.3

[1] $-\log\mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n) + \sum_j \log(c_j)$ [3] $-\frac{p}{2}\log(3) + \sum_i^p (\log(I_{ii})^{1/2} + \log(\lvert\hat{\theta}_i\rvert))$

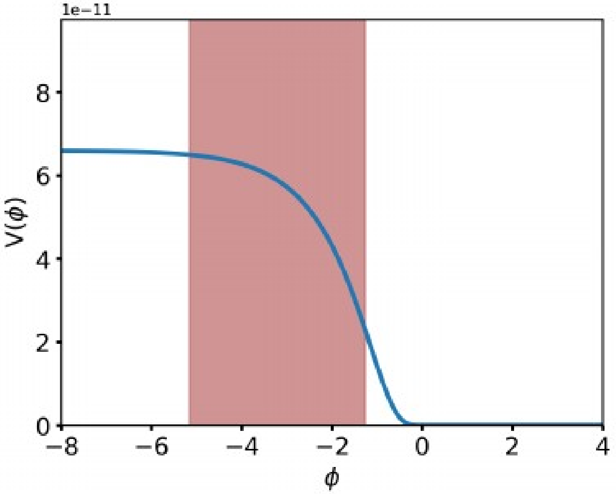


$$\theta_0\left(|\theta_1 + x|^{\theta_2} + x\right)$$

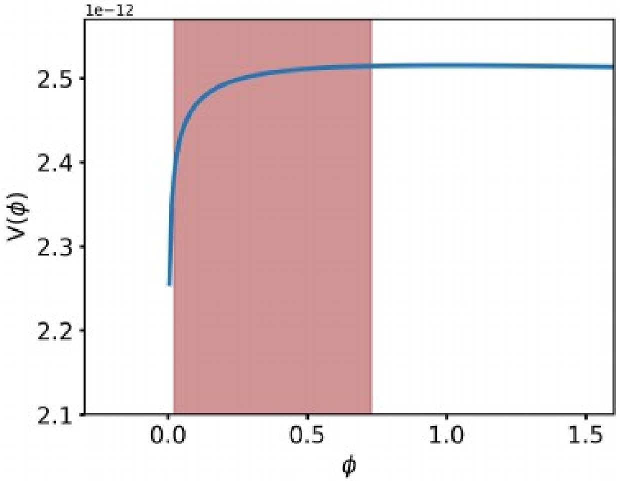
$$\theta_0 + \theta_1 x + \sqrt{x}$$

$$\theta_0 + \sqrt{x^2 + 2x}$$

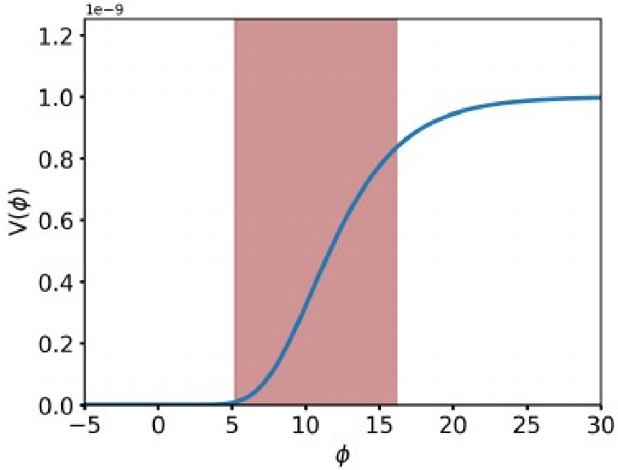
ESR readily Pareto-dominates all literature fits



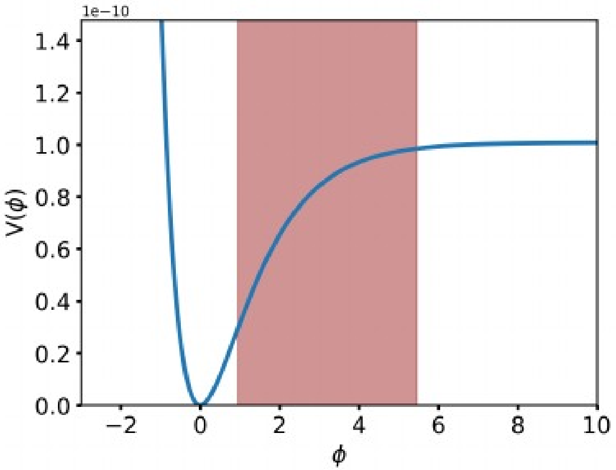
(a) $V(\phi) = e^{-e^{e^{e^{\phi}}}}$



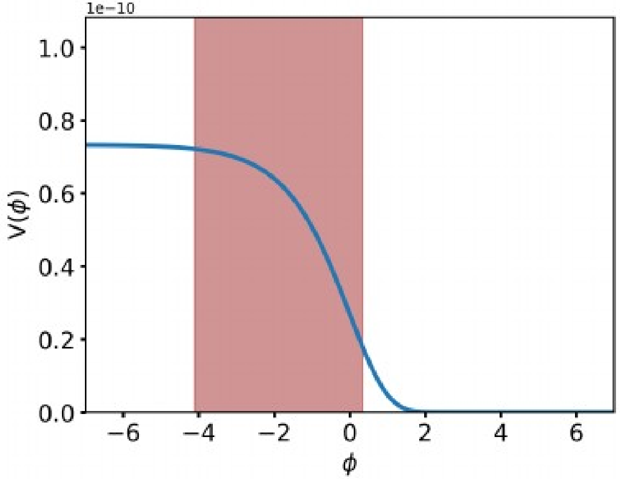
(b) $V(\phi) = \theta_0 \left(\theta_1 + \log(\phi)^2\right)$



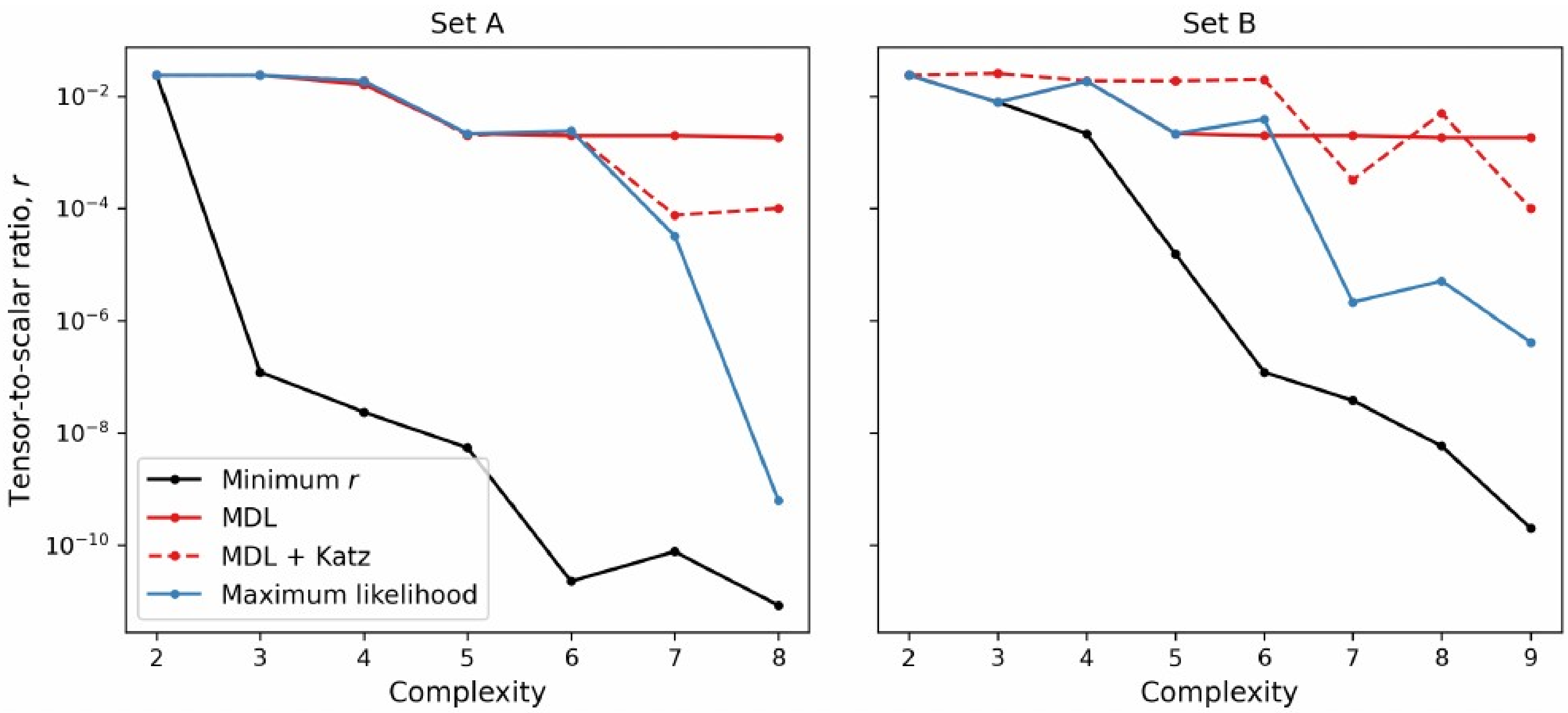
(c) $V(\phi) = |\theta_0|^{|\theta_1|^{\phi}}$



(d) Starobinsky



(e) $V(\phi) = \theta_0 e^{-e^{\phi}}$



Set A
Set B
Tensor-to-scalar ratio, r
10^-2
10^-4
10^-6
10^-8
10^-10
Complexity
2
3
4
5
6
7
8
9
Minimum r
MDL
MDL + Katz
Maximum likelihood

Set A

klog(n)

Rank	$V(\phi)$	Complexity	Prediction		Codelength			
			n_s	r	Residuals[1]	Function[2]	Parameter[3]	Total
1	$e^{-e^{e^{e^{\phi}}}}$	6	0.9678	0.002	-12.65	6.59	0.00	-6.06
2	$\theta_0 e^{-e^{\phi}}$	5	0.9668	0.002	-12.79	6.93	0.69	-5.16
3	$e^{\frac{1}{\sin(\sqrt{\phi})}}$	5	0.9634	0.012	-12.39	8.05	0.00	-4.35
4	$e^{-3e^{e^{\phi}}}$	6	0.9673	0.002	-12.38	8.32	0.00	-4.07
5	$\theta_0 - e^{3\phi}$	5	0.9670	2×10^{-4}	-12.77	8.05	0.69	-4.03
6	$\theta_0 - e^{2\phi}$	5	0.9673	5×10^{-4}	-12.74	8.05	0.69	-4.00
7	$\sin^2(\sin(\sin(\sqrt{\phi})))$	6	0.9619	0.013	-12.06	8.32	0.00	-3.74
8	$e^{e^{\phi}-e^{e^{e^{\phi}}}}$	8	0.9685	0.002	-12.50	8.79	0.00	-3.71
9	$\theta_0 e^{-e^{e^{\phi}}}$	6	0.9680	0.002	-12.62	8.32	0.69	-3.61
10	$(\theta_0 + e^{\phi})^2$	5	0.9676	0.002	-12.68	8.05	1.06	-3.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
—	$\theta_0\left(1 - e^{-\sqrt{2/3}\phi}\right)^2$	9	0.9678	0.003	-12.63	17.51	0.69	5.57
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24273	$\theta_0 \phi^2$	4	0.9669	0.132	31.82	5.55	0.69	38.06
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
38717	$\theta_0 \phi^4$	5	0.9508	0.262	168.23	6.93	0.69	175.85

[1] $-\log \mathcal{L}(\hat{\boldsymbol{\theta}})$ [2] $k\log(n) + \sum_j \log(c_j)$ [3] $-\frac{p}{2}\log(3) + \sum_i^p \left(\log(I_{ii})^{1/2} + \log\left(|\hat{\theta}_i|\right)\right)$

Set B

klog(n)

Rank	$V(\phi)$	Complexity	Prediction		Codelength			
			n_s	r	Residuals	Function	Parameter	Total
1	$e^{-e^{e^{e^{\phi}}}}$	6	0.9678	0.002	-12.65	6.59	0.00	-6.06
2	$\theta_0 e^{-e^{\phi}}$	5	0.9668	0.002	-12.79	6.93	0.69	-5.16
3	$\|\theta_0\|^{e^{e^{\phi}}}$	5	0.9674	0.002	-12.72	6.93	0.69	-5.09
4	$e^{e^{\phi}-e^{e^{\phi}}}$	8	0.9685	0.002	-12.50	8.79	0.00	-3.71
5	$\theta_0 e^{-e^{e^{\phi}}}$	6	0.9680	0.002	-12.62	8.32	0.69	-3.61
6	$e^{-e^{\frac{1}{\phi}}}$	5	0.9768	0.008	-8.65	5.49	0.00	-3.16
7	$e^{\theta_0 e^{e^{\phi}}}$	6	0.9674	0.002	-12.72	8.32	1.36	-3.04
8	$\|\theta_0\|^{e^{e^{e^{\phi}}}}$	6	0.9678	0.002	-12.65	8.32	1.39	-2.94
9	$e^{e^{\phi}} e^{-e^{e^{e^{\phi}}}}$	9	0.9685	0.002	-12.50	9.89	0.00	-2.62
10	$\|\theta_0\|^{\frac{1}{\phi}}$	4	0.9756	0.019	-8.77	5.55	0.69	-2.53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1272	$\theta_0\left(1-e^{-\sqrt{2/3}\phi}\right)^2$	9	0.9678	0.003	-12.63	17.51	0.69	5.57
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8697	$\theta_0\phi^2$	4	0.9669	0.132	31.82	5.55	0.69	38.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10839	$\theta_0\phi^4$	5	0.9508	0.262	168.23	6.93	0.69	178.81

Set A

Katz

Rank	$V(\phi)$	Complexity	Prediction		Codelength			
			n_s	r	Residuals	Function	Parameter	Total
1	$\theta_0\left(\theta_1+\log(\phi)^2\right)$	7	0.9649	8×10^{-5}	-12.90	9.28	1.39	-2.23
2	$\frac{1}{\left(\theta_0+e^{\phi}\right)^2}$	6	0.9654	0.002	-12.88	10.31	1.06	-1.51
3	$e^{\frac{\theta_0}{\phi}}$	4	0.9756	0.019	-8.77	6.98	0.69	-1.10
4	$\theta_0\left(\theta_1-e^{\phi}\right)^2$	7	0.9676	0.002	-12.68	10.77	1.39	-0.53
5	$\frac{\theta_0}{\phi+\log\left(\frac{\lvert\theta_1\rvert}{\phi}\right)}$	8	0.9649	1×10^{-4}	-12.90	11.06	1.39	-0.45
6	$\left(\theta_0-e^{\phi}\right)^2$	5	0.9676	0.002	-12.68	11.59	1.06	-0.04
7	$\theta_0\log(\phi)$	4	0.9783	0.024	-6.39	6.10	0.69	0.40
8	$\frac{1}{\theta_0+e^{\frac{\phi}{\theta_1}}}$	7	0.9649	0.001	-12.90	11.23	2.08	0.42
9	$\theta_0-\frac{\theta_1}{\phi}$	5	0.9768	0.010	-8.62	8.12	1.65	1.16
10	$\theta_0 e^{\theta_1 e^{\phi}}$	7	0.9668	0.002	-12.79	12.74	1.39	1.34
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16653	$\theta_0\phi^2$	4	0.9669	0.132	31.82	5.49	0.69	38.01
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
37635	$\theta_0\phi^4$	5	0.9508	0.262	168.23	7.15	0.69	176.08
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
—	$\theta_0\left(1-e^{-\sqrt{2/3}\phi}\right)^2$	9	0.9678	0.003	-12.63	11.42	0.69	-0.52

Set B

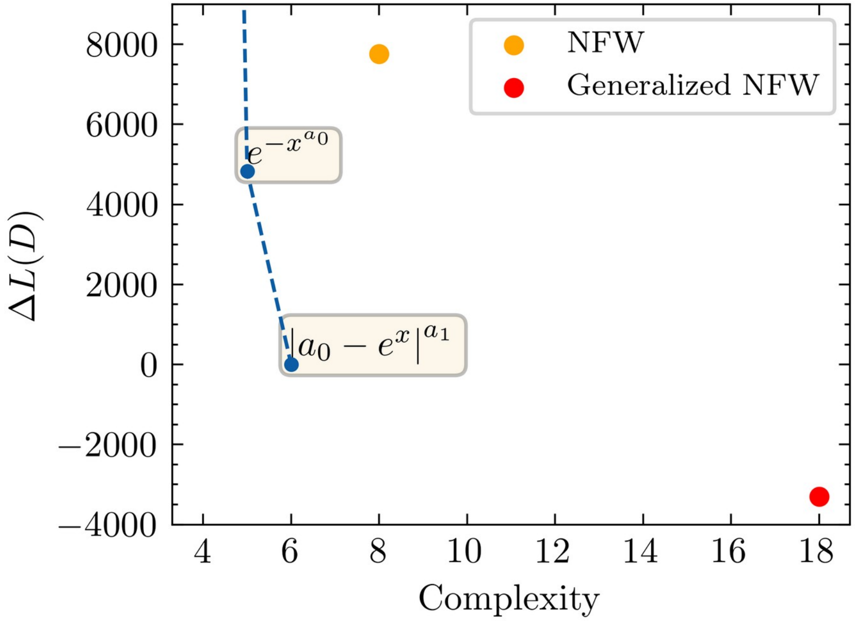
Katz

Rank	$V(\phi)$	Complexity	Prediction		Codelength			
			n_s	r	Residuals	Function	Parameter	Total
1	$\theta_0 \phi^{\frac{\theta_1}{\phi}}$	7	0.9649	3×10^{-4}	-12.90	8.92	1.39	-2.59
2	$\theta_0 \left(\theta_1 + \phi^{\phi}\right)$	7	0.9649	6×10^{-4}	-12.90	10.12	1.39	-1.39
3	$\theta_0 \phi^{\phi^{\theta_1}}$	7	0.9649	0.002	-12.89	9.55	2.71	-0.63
4	$\phi^{\theta_0} e^{\frac{\theta_1}{\phi}}$	8	0.9647	0.005	-12.83	9.37	2.83	-0.62
5	$\left(\phi + e^{\frac{\theta_0}{\phi}}\right)^{\theta_1}$	8	0.9636	0.012	-12.48	10.05	1.83	-0.60
6	$\lvert\theta_0\rvert^{\frac{\theta_1}{\phi}}$	5	0.9756	0.019	-8.77	6.88	1.39	-0.50
7	$e^{\theta_0 \lvert\theta_1\rvert^{\phi}}$	8	0.9653	0.020	-11.66	8.11	3.26	-0.30
8	$e^{\frac{\theta_0}{\phi}}$	4	0.9756	0.019	-8.77	8.04	0.69	-0.03
9	$\theta_0 e^{\frac{\theta_1}{\phi}}$	6	0.9653	0.020	-8.82	7.84	1.39	0.40
10	$\theta_0 \log(\phi)$	4	0.9783	0.024	-6.39	6.26	0.69	0.57
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12	$\theta_0 \left(1 - e^{-\sqrt{2/3}\phi}\right)^2$	9	0.9678	0.003	-12.63	12.64	0.69	0.70
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5401	$\theta_0 \phi^2$	4	0.9669	0.132	31.82	4.67	0.69	38.03
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8938	$\theta_0 \phi^4$	5	0.9508	0.262	168.23	4.67	0.69	180.84

Inferring halo density profiles

Richard Stiskalek

From simulations



From observations

Alicia Martin

F568-3